A COMPUTER BASED VERSATILE METHOD FOR IDENTIFYING PROTEIN CODING DNA SEQUENCES USEFUL AS DRUG TARGETS

Field of the present invention

5   The present invention relates to a versatile method of identifying genes having invariant peptides as functional signatures in a genome especially in SARS using software GeneDecipher. Further, it relates to a four novel genes of SARS and their corresponding proteins. Lastly, it also relates to a method of drug target development in the management in a disease condition

10   **Background and prior art references of the present invention**

The most reliable way to identify a protein coding DNA sequence (gene) in a newly sequenced genome is to find a close homolog from other organisms (BLAST (Altschul,S.F et al., 1990) and FASTA (Pearson,W.R., 1995)). Four nucleotides in a DNA sequence are not randomly distributed. The statistical distribution of nucleotides within a coding region

15   is significantly different from the non-coding (Bird,A., 1987). Methods based on Hidden Markov Models (HMM) have used these statistical properties most efficiently (Salzberg,S.L et al., 1998; Delcher,A.L et al.,1999; Lukashin,A.V. and Borodovsky,M., 1998) and are able to predict ~97-98 % of all the genes in a genome when compared with published annotations (Delcher,A.L et al., 1999). Using HMM, various algorithms like

20   GeneMark, Glimmer etc. have been developed to predict genes in prokaryotes. Glimmer 2.0 is the most successful method among all existing methods (Delcher,A.L et al., 1999). However, Glimmer also predicts 7-20% additional genes (false positives).

Each gene prediction method has its own strengths and weaknesses (Mathe,C. et al., 2002). Since the prediction is usually dependent on the training set, shortcomings arise because

25   statistics for a coding region vary across various genomes. Also, these methods are unable to efficiently predict genes small in length (< 100 amino acids), because it's very difficult to detect these genes by similarity searches or by statistical analysis. The problem becomes more severe in case of horizontal gene transfer (Kehoe,M.A et al.,1996). In this case statistical distribution of the nucleotide sequence of these genes differs within a genome

30   itself.

1

The said method of the invention is based upon the observation that the difference between total number of theoretically possible peptides of a given length and that which are actually observed in nature, increases drastically as this length of peptide increases. For example, only about 2% of the theoretically possible heptapeptides are observed in a pool of 56 completely sequenced prokaryotic genomes. At octapeptide level this number reduces to less than 0.1%. Moreover, it is interesting to note that most of these peptides selected by nature are found only in the coding regions and very rarely in theoretically translated non-coding regions. This observation has prompted us to exploit this exclusivity of natural selection of peptides that are present in protein coding sequences to differentiate between coding and non-coding regions.

In principle, using longer peptides to score a query ORF is always preferable to using shorter ones (Salzberg,S.L. et al., 1998), but only if sufficient data is available to estimate statistical parameters required to train the prediction algorithm. In case we use peptides of length 8 or more amino acids, it is difficult to get sufficient data to estimate the training parameters. This is because likelihood of an octapeptide being shared between two polypeptides is less than that of a heptapeptide. So we consider the length of 7 amino acids as optimum for scoring of an ORF.

The novelty of the said method is that it works on the basis of protein coding sequences at amino acid, not at nucleotide sequence level. It is noteworthy that the method does not need an organism specific training set, which is an obvious advantage over other methods. Unlike other methods, GeneDecipher does not employ any landmarks like ribosome binding sites, promoter sequences, transcription start sites or codon usage biases to predict the coding genes and their start locations. In addition, this method overcomes the difficulties of gene prediction for smaller genomes (Chen,L et.al., 2003) like SARS-CoV. Other than gene prediction, this method can also be utilized for similarity searches for polypeptides, putative functional assignment to proteins (based on presence of the oligo-peptide motifs), and in phylogenetic domain analysis, indicating the generic-ness and versatility of the method.

Severe acute respiratory syndrome (SARS) has emerged as a life threatening disease. Early reports on SARS appeared from China (Ksiazek *et al.,* 2003) and subsequently, cases of

2

SARS were reported from Taiwan, Vietnam, Canada, Singapore and other countries. The range of symptoms observed in SARS affected patients are fever, dry cough, dyspnea, headache, and hypoxemia. Typical laboratory findings include lymphopenia and mildly elevated aminotransferase levels. Death may result from progressive respiratory failure due to alveolar damage (Tsang *et al.*, 2003). On an average, the mortality rate was 4%, though it varied widely according to the geographic location (WHO report, 2003) and with the strain implicated. SARS isolates from different parts of the world have been sequenced recently. Sequence analysis of nucleic acid fragments isolated from cytopathic Vero cell cultures showed that the, encoded protein sequences were similar to protein of other coronaviruses (Drosten *et at.*, 2003). However, at the nucleic acid level, no similarity was observed with any sequence in the database indicating substantial diversity. Phylogenetic analysis showed that the isolated sequence is distinct and is placed between groupZ and group3 coronavimses in the tree (Marra et al., 2003).

Current computational methods like GeneMark.hmm (Lukashin and Borodovsky, 1998), Glimmer (Salzberg *et al.*, 1998), etc. face difficulty in analyzing the SARS genome due to its small size. Methods based on Hidden Markov Models (HMM) require thousands of parameters for training. This makes these methods less suitable for analyzing smaller genomes. The problem compounds in the case of SARS-CoV genomes which are about 30kb_Jn length. Even the method most suitable for viral gene prediction till date ZCURVE_CoV (Chen et al., 2003) needs 33 parameters for training.

Objects of the present invention

The main object of the present invention is to provide a computer based method for predicting protein coding DNA sequences (genes) useful as drug targets.

Another main object of the present invention is to develop a versatile method of identifying genes having invariant peptides as functional signatures in a genome using software GeneDecipher.

Yet another object of the present invention is to develop a method of identifying genes having functional signatures in the SARS virus.

Still another object of the present invention is to identify novel genes from the SARS genome.

Still another object of the present invention is to develop a novel peptides corresponding to the novel genes of the SARS.

Still another object of the present invention is to develop a method of drug development in the management in a disease condition.

Still another object of the present invention is to develop a method of drug development in the management of SARS virus.

Still another object of the present invention is to develop a microprocessor-based system for performing the aforementioned methods.

Still another object of the present invention is to develop a computer based system for performing the aforementioned methods.

Still another object of the present invention relates to a novel computer based method for performing genome-wise comparison of several organisms.

Yet another object of the present invention is to develop a method useful for identification of novel protein coding DNA sequences useful as potential drug targets and can serve as drug screen for broad spectrum antibacterial as well as for specific diagnosis of infection.

Still another object of the present invention is to identify strain specific or organism specific protein coding genes.

Yet another object of the method of invention is to identify protein coding DNA sequences (exons) in eukaryotic organisms.

Another object of the present invention is to assignment of function to hypothetical Open Reading Frames (proteins) of unknown function through exact amino acid sequence identity signature.

Summary of the present invention

The present invention relates to a versatile method of identifying genes having invariant peptides as functional signatures in a genome especially in SARS using software GeneDecipher, said method comprising steps of generating peptide libraries from the known genomes with peptide of length 'N' computationally arranged in an alphabetical order, artificially translating the test genome to obtain peptide, identifying the reading frames in the peptide on the basis of overlappings with the peptide libraries, converting each peptide sequence into an alphanumeric sequence with one corresponding to each

4

reading frame, training Artificial Neural Network (ANN) with sigmoidal learning function to the alphanumeric sequence, deciphering the protein coding regions in the test genome, thus, identifying invariant peptides serving as functional signatures, also, four novel genes of SARS and their corresponding proteins, lastly, a method of drug target development in the management in a disease condition

Detailed description of the present invention

Accordingly, the present invention relates to a versatile method of identifying genes having invariant peptides as functional signatures in a genome especially in SARS using software GeneDecipher, said method comprising steps of generating peptide libraries from the known genomes with peptide of length 'N' computationally arranged in an alphabetical order, artificially translating the test genome to obtain peptide, identifying the reading frames in the peptide on the basis of overlappings with the peptide libraries, converting each peptide sequence into an alphanumeric sequence with one corresponding to each reading frame, training Artificial Neural Network (ANN) with sigmoidal learning function to the alphanumeric sequence, deciphering the protein coding regions in the test genome, thus, identifying invariant peptides serving as functional signatures, also, four novel genes of SARS and their corresponding proteins, lastly, a method of drug target development in the management in a disease condition

A versatile method of identifying genes having invariant peptides as functional signatures in a genome using software GeneDecipher, said method comprising steps of:

    a. generating peptide libraries from the known genomes with peptide of length 'N' computationally arranged in an alphabetical order,

    b. artificially translating the test genome to obtain peptide,

    c. identifying the reading frames in the peptide on the basis of overlappings with the peptide libraries,

    d. converting each peptide sequence into an alphanumeric sequence with one corresponding to each reading frame,

    e. training Artificial Neural Network (ANN) with sigmoidal learning function to the alphanumeric sequence,

    f. deciphering the protein coding regions in the test genome, and

g. identifying invariant peptides serving as functional signatures.

In yet another embodiment of the present invention the ANN has one or more input layer, one or more hidden layer with varying number of neurons, and one or more output layer.

In still another embodiment of the present invention the number of neurons is preferably 30.

In yet another embodiment of the present invention the length of the 'N' is 4 or more.

In yet another embodiment of the present invention the sigmoidal learning function has five parameters comprising total score, mean, fraction of zeroes, maximum continuous non-zero stretch, and variance.

One more embodiment of the present invention a method of identifying genes having functional signatures in the SARS virus, said method comprising steps of:

a) generating heptapeptide libraries of non-SARS virus genomes with peptide of length 'N' computationally arranged in an alphabetical order,

b) artificially translating the SARS virus genome to obtain peptide,

c) identifying the six reading frames in the peptide on the basis of overlappings with the heptapeptide libraries,

d) converting each peptide sequence into an alphanumeric sequence with one corresponding to each reading frame,

e) training Artificial Neural Network (ANN) with sigmoidal learning function to the alphanumeric sequence,

f) deciphering the protein coding regions in the SARS virus genome, and

g) identifying invariant peptides of SARS virus serving as functional signatures.

In yet another embodiment of the present invention the method discloses 15 protein-coding regions.

In still another embodiment of the present invention the method identifies four novel genes SARS174, SARS68, SARS61, and SARS90.

In yet another embodiment of the present invention the ANN has one or more input layer, one or more hidden layer with varying number of neurons, and one or more output layer.

6

In still another embodiment of the present invention the number of neurons is preferably 30.

In yet another embodiment of the present invention the length of the 'N' is 4 or more.

In still another embodiment of the present invention the sigmoidal learning function has five parameters comprising total score, mean, fraction of zeroes, maximum continuous non-zero stretch, and variance.

In yet another embodiment of the present invention the method is better than the conventional methods.

In still another embodiment of the present invention a Sars174 gene of SARS virus of SEQ ID No. 1.

SEQ ID No. 1 is as given below:

GTGACGAGCTTGGCACTGATCCCATTGAAGATTATGAACAAAACTGGAACACTAAGCATGGCA
GTGGTGCACTCCGTGAACTCACTCGTGAGCTCAATGGAGGTGCAGTCACTCGCTATGTCGAC
AACAATTTCTGTGGCCCAGATGGGTACCCTCTTGATTGCATCAAAGATTTTCTCGCACGCGCG
GGCAAGTCAATGTGCACTCTTTCCGAACAACTTGATTACATCGAGTCGAAGAGAGGTGTCTAC
TGCTGCCGTGACCATGAGCATGAAATTGCCTGGTTCACTGAGCGCTCTGATAAGAGCTACGA
GCACCAGACACCCTTCGAAATTAAGAGTGCCAAGAAATTTGACACTTTCAAAGGGGAATGCCC
AAAGTTTGTGTTTCCTCTTAACTCAAAAGTCAAAGTCATTCAACCACGTGTTGAAAAGAAAAAG
ACTGAGGGTTTCATGGGGCGTATACGCTCTGTGTACCCTGTTGCATCTCCACAGGAGTGTAAC
AATATGCACTTGTCTACCTTGA

In yet another embodiment of the present invention a Sars gene as claimed in claim 14, wherein the length of the gene is 525 bp.

In still another embodiment of the present invention a Sars174 protein of SARS virus of SEQ ID No. 2.

SEQ ID No. 2 is as given below:

VTSLALIPLKIMNKTGTLSMAVVHSVNSLVSSMEVQSLAMSTTISVAQMGTLLIASKIFSHARASQCA
LFPNNLITSSRREVSTAAVTMSMKLPGSLSALIRATSTRHPSKLRVPRNLTLSKGNAQSLCFLLTQK
SKSFNHVLKRKRLRVSWGVYALCTLLHLHRSVTICTCLP*

In yet another embodiment of the present invention a Sars174 protein as claimed in claim 16, wherein the length of the protein is 174 aa.

In still another embodiment of the present invention A Sars68 gene of SARS virus of SEQ ID No. 3.

SEQ ID No. 3 is as given below:

TTGGACCTGAGCATAGTGTTGCAGATTATCACAACCACTCAAACATTGAAACTCGACTCCGCA
AGGGAGGTAGGACTAGATGTTTTGGAGGCTGTGTGTTTGCCTATGTTGGCTGCTATAATAAGC
GTGCCTACTGGGTTCCTCGTGCTAGTGCTGATATTGGCTCAGGCCATACTGGCATTACTGGTG
ACAATGTGGAGACCTTGA

In still another embodiment of the present invention A Sars gene as claimed in claim 18, wherein the length of the gene is 207 bp.

In yet another embodiment of the present invention A Sars68 protein of SARS virus of SEQ ID No. 4.

SEQ ID No. 4 is as given below:

LDLSIVLQIITTTQTLKLDSAREVGLDVLEAVCLPMLAAIISVPTGFLVLVLILAQAILALLVTMWR*

In still another embodiment of the present invention A Sars68 protein as claimed in claim 20, wherein the length of the protein is 68 aa.

In yet another embodiment of the present invention A Sars61 gene of SARS virus of SEQ ID No. 5.

SEQ ID No. 5 is as given below:

ATGGTGACTTCTTGCATTTTCTACCTCGTGTTTTTAGTGCTGTTGGCAACATTTGCTACACACC
TTCCAAACTCATTGAGTATAGTGATTTTGCTACCTCTGCTTGCGTTCTTGCTGCTGAGTGTACA
ATTTTTAAGGATGCTATGGGCAAACCTGTGCCATATTGTTATGACACTAATTTGCTAG

In still another embodiment of the present invention A Sars gene as claimed in claim 22, wherein the length of the gene is 186 bp.

In yet another embodiment of the present invention A Sars61 protein of SARS virus of SEQ ID No. 6.

SEQ ID No. 6 is as given below:

MVTSCIFYLVFLVLLATFATHLPNSLSIVILLPLLAFLLLSVQFLRMLWANLCHIVMTLIC*

In still another embodiment of the present invention A Sars61 protein as claimed in claim 24, wherein the length of the protein is 61 aa.

Another embodiment of the present invention a method of drug development in the management in a disease condition, said method comprising step of using a proposed drug for blocking the functioning of one or more invariant peptides as functional signatures identified by the instant method.

Further embodiment of the present invention a method of drug development in the management of SARS virus, said method comprising step of using a proposed drug for blocking the functioning of one or more invariant peptides as functional signatures selected from a group comprising Sars174, Sars68, Sars61, and Sars90.

In yet another embodiment of the present invention the Sars174 is involved in ABC transporter ATP binding protein.

In still another embodiment of the present invention the Sars68 is a major facilitator superfamily protein.

In yet another embodiment of the present invention the Sars90 is involved in NADH Dehydrogenase I chain.

The present invention relates to a microprocessor based system for performing the methods of the invention which comprises:

      i)   means of determining the amino acid sequence window for creation of peptide library and subsequent origin tagging,

      ii)   means of comparing the peptide library,

      iii)   locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location, and

      iv)   joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

A computer based system for performing the methods of the invention further comprising a central processing unit, executing peptide library creating program (PEPLIB), peptide library matching program (PEPLIMP), peptide stitching program (PEPSTITCH), peptide extraction program (PEPXTRACT) wherein the said programs are all stored in a memory device accessed by the central processing unit connected to a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs with a user interface device.

The present invention relates to a novel computer based method for predicting protein coding DNA sequence useful as drug targets, the said computational method involves creation of peptide libraries from protein sequences of several organisms and subsequent comparison leading to identification of protein coding DNA sequences, and to this end

9

several protein coding DNA sequences (genes) have been identified by this novel computer based method. The invention relates to a novel method of converting DNA sequence to alphanumeric sequence by the use of peptide library and the invention also provides a method for use of artificial neural network (feed forward back propagation topology) with one input layer, one hidden layer with 30 neurons and one output layer for identification protein coding DNA sequences. The invention further relates to a method for training of neural networks using sigmoid as a learning function with five parameters namely total score, mean, fraction of zeroes, maximum continuous non-zero stretch and variance for identification of protein coding DNA sequence and the present method is useful for identification of new protein coding regions which can serve as drug screen for broad-spectrum antibacterials as well as for specific diagnosis of infections, and in addition, for assignment of function to newly identified proteins of yet unknown functions. The method allows identification of species or strain specific protein coding genes. This method also can be extended to any protein coding sequence identification even in eukaryotic genomes. This invention relates to a computer-based method for predicting protein coding DNA sequences useful as drug targets. More particularly this invention relates to a method for identification of novel genes in genome sequence data of various organisms, useful as potential drug targets. This invention further provides a method for assignment of function to hypothetical Open Reading Frames (proteins) of unknown function through exact amino acid sequence identity signature.

Emergence of high throughput sequencing technologies has necessitated identification of novel protein coding DNA sequences (genes) in newly sequenced genomes. The invention provides a novel method of converting DNA sequence to alphanumeric sequence by the use of peptide library. The invention also provides a method for use of artificial neural network (feed forward back propagation topology) with one input layer, one hidden layer with 30 neurons and one output layer for identification protein coding DNA sequences. The invention further provides a method for training of neural networks using sigmoid as a learning function with five parameters namely total score, mean, fraction of zeroes, maximum continuous non-zero stretch and variance for identification of protein coding DNA sequence.

10

The applicants have invented a novel computer based method to identify protein coding DNA sequences by comparing with peptide library containing millions of peptides obtained from protein sequences of many organisms that has withstood natural selection. The method describes a generic and versatile new approach for gene identification. The computational method determines gene candidates among all possible Open Reading Frames (ORF) of a given DNA sequence through the use of a peptide library and an artificial neural network. The peptide library consists of all possible overlapping heptapeptides derived from proteins of completely sequenced 56 prokaryotic genomes. A given query ORF qualifies as a gene based upon the abundance and distribution pattern of library heptapeptides (heptapeptides present in library) along the ORF. Performance of the method is characterized by simultaneous high values of sensitivity and specificity. An analysis of 10 completely sequenced prokaryotic genomes is provided to demonstrate the capabilities of the method of the invention.

The present method also allows prediction of alternate target against a specific peptide motif of a pathogenic organism or any host protein target responsible for a disease process. The method could be extended with different peptide lengths to obtain larger number of protein coding genes and also for eukaryotes and multicellular organisms.

Other and further aspects, features and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention given for the purpose of disclosures.

Accordingly the invention provides a computer-based method for predicting protein coding DNA sequences useful as drug targets wherein the said method comprises the steps of:

i) generating computationally overlapping peptide libraries from all the protein sequences of the.

selected organisms available at http://www.ncbi.nlm.nih.gov,

ii) sorting computationally the peptides of length 'N' obtained as above, alphabetically, according to single letter amino acid code,

iii) cataloging every peptide and their unique occurrence different organisms,

iv) converting DNA sequence to alphanumeric sequence using peptide library obtained from steps 1 and 2,

v) retrieving all possible open reading frames (ORFs) from the alphanumeric sequence,

vi) training of the modified neural network for discriminating protein coding and non-coding DNA sequences,

vii) predicting DNA coding sequences in the open reading frames (obtained in step 4) using trained neural network,

viii) removing the encapsulated protein coding DNA sequences (genes within genes).

In an embodiment to the present invention the sliding peptide window of length 'N' may range from 4 to any length of amino acid residues.

In another embodiment to the present invention the conversion of the DNA sequence to alphanumeric sequence may be carried out computationally using characters selected from but not restricted to 's', '*', '-', (0-9).

In further embodiment the training of the modified neural network for discriminating protein coding form non-coding DNA sequences is done using parameters but not limited to these such as score, mean, fraction of zeros, maximum continuous non-zero stretch and variance.

In still another embodiment to the invention the modified neural network may consist of but not limited to one input layer, one hidden layer with 30 neurons and one output layer. The method may consist of multiple input, hidden and output layer with varying number of neurons at any layer.

In yet another embodiment of the present invention the peptide library data may be taken from any organism but not specifically limited to those used in the invention.

Brief description of the computer programs:

1. File Name: genedcodchr.cxx

Application: Translation of nucleotide sequence (FASTA file format) into 6 hypothetical polypeptides in 6 respective frames.

Input format : <Program_name> <Nucleotide_file> <Output1> <Output2> <frame>

e.g.,   ./genedcodchr ecoli.fna pf1 pr1 0

Output format:

AGTFYRYmGHVNMKIYTASLPTYRYGYFSHRED.....HGOIEKSDW EzDFGTRE

2. File Name: searchchr.cxx

Application: Converts the polypeptide file into an alphanumeric sequence through a heptapetide library (given as an input) search.

Input format :< Program_name> 7 <peptide library file name> out Y <Input1> <Input2> <Output1> <Output 2>

e.g.,   ./searchchr 7 ecoli.peplib out Y pf1 pr1 bf1 br1

Output format:

s11245000010900030000200000230000000000*******0001000..........

3. File Name: cutf.c

Application: Cuts all possible ORFs (i.e., all 's' to '*' regions) from the alphanumeric sequence of forward strand and generates a file containing locations of all the 's' in alphanumeric sequence.

Input format :< Program_name> <Input file name> <Output1> <Output2>

e.g.   ./cutf bf1 unknown_bf1 bf1_location

Output format: output1- s1111000s00000000563*, output2- starting locations of 's' in a column.

4. File Name: cutr.c

Application: Cuts the all possible ORFs (all 's' to '* regions) from the reverse strand's alphanumeric sequences and produces a file which contains the starting locations in alphanumeric sequence file for all 3 forward frames corresponding to all ORFs.

Input format :< Program_name> <Input file name> <Output1> <Output2>

e.g.   ./cutr br1 unknown_br1 br1_location

Outputformat: output1-*0103400002222000679000000s000001000200s00230000s, output2- starting location of 's'

5. File Name: stat.c

Application: Calculates the five parameters: fraction of zeros, mean, total score, length of maximum continuous stretch, and variance for a given alphanumeric sequence.

Input format :< Program_name> <Input file name><Output> 1

e.g.   ./stat unknown_bf1 bf1.data 1

Output format: 0.334   3.2   48   15   0.452 1

13

6. File Name: train .c

Application: Training of Artificial Neural Network (single hidden layer, 1 input and 1 output layer) with feed forward back propagation algorithm and using sigmoid (= 1) as a learning function.

Input format :< Program_name> <Input specification file name> <Input1> <Input2> <Input3> > output

e.g. ./train train.spec.fast trainset.data validateset.data testset.data > train.net

Output format: output containing the final neural network wieghts in a single column.

7. File Name: recognize.c

Application: Recognizes a given pattern on the basis of trained weights and generates a probability value as output.

Input format :< Program_name> <Input specification file name> <Input1> <Input2> <Output>

e.g. ./recognize recognize.spec bf1.data train.net f1.out

Output format: pat1   probability   <value>

8. File Name: Filter_prediction.c

Application: Filters out the completely overlapping ORFs in same frame based on probability and length parameter.

Input format :< Program_name> <Input1> <Input2> <Output>

e.g. ./Filter_prediction f1.out unknown_bf1 bf1.out.res

Output format: pat1   probability   <value>   <integer string>

9. File Name: locationf.c

Application: Filters out the genes of length <20 amino acids, and reports starting location of the remaining ones with the alphanumeric sequence for all 3 forward frames.

Input format :< Program_name> <Input1> <Output> <Input2>

e.g. · ./locationf bf1.out.res bf1.out.res1 bf1_location

Output format:<Pattern No> <Probability value> <integer string> <Start> <End>

10. File Name: locationr.c

Application: Filters out the genes of length <20 amino acids, and reports starting location of the remaining ones with the alphanumeric sequence for all 3 reverse frames.

14

Input format :< Program_name> <Input1> <Output> <Input2>

e.g.   ./locationr br1.out.res br1.out.res1 br1_location

Output format:<Pattern No> <Probability value> <integer string> <Start> <End>

11. File Name: finalf.c

Application: Converts the start and end locations of the alphanumeric sequence into the corresponding genome locations for 3 forward frames.

Input format :< Program_name> <Input1> <Input2> <Input3> <Output>

e.g.   ./finalf bf1.out.res1 bf2.out.res1 bf3.out.res1 Final_outputf

Output format:<Start> <End> <frame> <length> <Probability value> <integer string>

12. File Name: finalr.c

Application: Converts the start and end locations of the alphanumeric sequence into the corresponding genome locations for 3 reverse frames.

Input format :< Program_name> <Input1> <Input2> <Input3> <Output>

e.g.   ./finalf br1.out.res1 br2.out.res1 br3.out.res1 Final_outputr

Output format:<Start> <End> <frame> <length> <Probability value> <integer string>

13. File Name: sort.c

File Name: sort.c

Applications: Prints the finally predicted genes into descending order along the genome start location.

Input format :< Program_name> <Input1> <Input2> <Input3> <Output>

e.g.   ./sort   Final_outputf   Final_outputr   OUTPUTF_with_encap OUTPUTR_with_encap OUTPUT

Output format:<Start> <End> <Probability value>

14. File Name: removeencap.c

Application: Removes encapsulated genes found in other five frames.

Input format :< Program_name> <Input1> <Input2> <Input3> <Output>

e.g.   ./removeencap OUTPUTF_with_encap OUTPUTR_with_encap OUTPUT OUTPUTF OUTPUTR

Output format:<Start> <End> <frame> <length> <Probability value> <integer string>

The present invention relates to a novel computer based method for predicting protein coding DNA sequences useful as drug targets. In this method occurrence of oligopeptide signatures have been used as probes. The method is versatile and does not necessarily require organism specific training set for the Artificial Neural Network. The method is not only dependent on statistical analysis but also integrates with the biological information that is retained in the conserved peptides, which withstood evolutionary pressure. Logical extension of the method will be to predict protein coding DNA sequences (exons) in eukaryotic genomes.

Brief description of the accompanying drawings

Figure 1 shows a logic circuit of GeneDecipher.

Figure 2 shows a architecture of neural network.

Figure 3 shows analysis of results of GeneDecipher on 10 organisms.

The method has been described in five major steps (as shown in Figure1):

1. Generation of a peptide library
2. Artificial translation of a given genome into 6 reading frames
3. Conversion of each translated sequence into an alphanumeric sequence. (one corresponding to each reading frame)
4. Training of artificial neural network (ANN).
5. Deciphering genes using trained ANN.

1. Generation of peptide library

The method requires a reference peptide library to predict genes in a given genome. In the present invention, the applicants have used proteins from 56 completely sequenced prokaryotic genomes. The protein files for our database were obtained in FASTA format from ftp://ftp.ncbi.nlm.nih.gov/genomes. To prepare a peptide library for deciphering genes in a particular genome, the applicants exclude protein file(s) belonging to that particular species from our database in order to avoid any bias. For example, when analyzing *E.coli*-k12 genome the protein files corresponding to all strains of *E.coli* were excluded from the database to create the peptide library. This has been done to eliminate the signal that is obtained from peptides of that organism, which would be the case while

analyzing a newly sequenced genome. This strengthens the method in terms of gene prediction on a newly sequenced genome for which annotated protein file is not available.

While creating peptide library all possible overlapping heptapeptides have been taken care of by shifting the window by one amino acid. Redundant peptides were eliminated from the peptide library and each peptide is given an occurrence value based on number of discrete organisms in which it is present.

This occurrence value is a measure of conservation of a heptapetide in coding regions. Presence of a heptapeptide with high occurrence value in an ORF increases the likelihood of that ORF being a protein coding gene. In our algorithm, occurrence value of 9 or more is treated as 9 based on the assumption that if a heptapeptide is present in 9 or more than 9 different organisms' protein files, it can be considered as highly conserved heptapeptide. It is not worthwhile to use any higher value to further discriminate the amount of conservation.

The heptapeptide library database consists of two columns, first for heptapeptide sequence and second for score (occurrence value) of that heptapeptide. Heptapeptides are sorted in dictionary order. The peptide library database also retains other information about the heptapeptides, like the accession number and NCBI annotation of all proteins containing the particular heptapeptide. This can be utilized for putative function prediction of a given ORF. Same approach can be used for phylogenetic domain analysis also.

2. Artificial translation of a given genome into 6 reading frames

Second step in the algorithm is artificial translation of the whole query genome in all six reading frames using a standard codon table. However user specified codon table may be used wherever necessary. Applicants used letter 'z' corresponding to the stop codons TTA, TAG and TGA, and letter 'b' for all triplets containing any non standard nucleotide(s) (K, N, W, R, and S etc.) while artificially translating the genome.

3. Conversion of each translated sequence into an alphanumeric sequence (one corresponding to each reading frame)

The next step in our algorithm is to convert artificially translated amino acid sequence with stop codon (z) interruption, into an alphanumeric sequence. Applicants search each overlapping heptapeptide in the peptide library, assign a corresponding number

(occurrence value), and append it to the alphanumeric sequence. If a heptapeptide is not present in the library Applicants assign the number 0. If a heptapeptide begins with an amino acid corresponding to any of the start codon ATG,GTG and TTG Applicants append character 's' in the alphanumeric sequence. This will be helpful to detect the location of a probable start codon. In case a heptapeptide contains character 'z' Applicants append a character '*' corresponding to that heptapeptide. Thus consecutive seven '*' (*******) in the alphanumeric sequence is a signal for stop codon. Applicants append '-' character for any heptapeptide containing character 'b'. This signals the presence of a non standard nucleotide character and conveys no information about sequence being a part of gene or non-gene. So, the alphanumeric sequence thus generated contain 13 characters viz. any integer (0-9), 's', '*', and '-'. In this way, Applicants convert all six translated protein files into six alphanumeric sequences.

4. Training of artificial neural network (ANN)

The neural network used here has a multi-layer feed-forward topology. It consists of one input layer, one hidden layer, and an output layer. This is a 'fully-connected' neural network where each neuron i is connected to each unit j of the next layer (Figure 2). The weight of each connection is denoted by w_{ij}. The state I_i of each neuron in the input layer is assigned directly from the input data, whereas the states of hidden layer neurons are computed by using the sigmoid function, $h_j = 1 / (1 + \exp - (w_{j0} + w_{ij} I_i))$, where, w_{j0} is the bias weight, and $= 1$.

The back propagation algorithm is used to minimize the differences between the computed output and the desired output. One thousand cycles (epochs) of iterations are performed. Subsequently, the epoch with minimum error in validation set is identified and the corresponding weights (w_{ij}) are assigned as the final weights for the ANN. The network trains on the training set, checks error and optimizes using the validation set through back propagation.

The 'training set' consists of 1610 E.coli-k12 NCBI listed protein coding genes and 3000 E.coli-k12 ORFs (a stretch of sequence of length more than 20 amino acids and having start codon, stop codon in the same frame) which have not been reported as genes (non-genes). The 'validation set' has 1000 known genes and 1000 non-genes from E.coli-k12,

18

distinct from those used in the training set. The 'test set' contains another 1000 genes and 1000 non-genes from the same organism. For training of the ANN, genes and the non-genes are assigned a probability value of 1 and 0 respectively.

To train the neural network, first Applicants convert all the *E.coli*-k12 genes and non-genes into corresponding alphanumeric strings by the method described above (steps 2 and 3). Here it is important to note that the alphanumeric sequences corresponding to a gene is number rich compared to the alphanumeric sequences corresponding to non-genes. To quantify this number richness of an alphanumeric sequence, five parameters derived from the alphanumeric sequence have been selected. These five parameters are as follows:

1. Total Score

This is an algebraic sum of all the integers of a given alphanumeric sequence. Here rule of thumb is higher the score, more are the chances to qualify as a gene.

2. Fraction of zeroes

Fraction of zeroes equals to total no. of zero characters in the alphanumeric sequence divided by total no. of characters in the sequence. More the fraction of zeros, lesser is the chance to qualify as a gene.

3. Mean

Mean equals to total score divided by total length of the sequence. Higher the Mean, more is the chance to qualify as a gene. Virtually this parameter seems same as a total score but it is important because this incorporates the length of the sequence also (score per unit length)

4. Variance

It is the variance of occurrence values about the mean occurrence value for the whole ORF.

5. Length of the maximum continuous non zero stretch

Higher the value of this parameter more is the chance to qualify as a gene. Consider a sequence region like '45'. Here, '4'denotes a heptapeptide conserved in 4 organisms, and the succeeding '5' denotes an overlapping heptapeptide conserved in 5 organisms. So if there exists at least one organism which is common between these two sets, eventually Applicants have an octapeptide common between that organism and the query ORF. This raises our confidence level in prediction of the coding region. For example, sequence

19

's45467000000*******' is more likely to be a gene when compared to sequence 's40540607000*******'. This is because there are greater chances of presence of conserved longer peptide in the first sequence. Value of the parameter is 5 for first string and 2 for second one. However, other parameters used in the algorithm can not discriminate between these two sequences.

While calculating these parameters from the alphanumeric sequences, characters such as 's', '*' and '-' have been excluded.

To find an optimum combination, the neural network is trained using all the five parameters together. Parameters corresponding to alphanumeric sequences of genes and non-genes are calculated. The training, validation and test sets contain 6 columns, first 5 columns contains values of the 5 parameters and the last column contains the number '1' for genes and the number '0' for non-genes.

The number of neurons in the input layer was equal to the number of input data points. The optimal number of neurons in the hidden layer was determined by hit and trial while minimizing the error at the best epoch for the network. Computer program to compute all 5 parameters and for the artificial neural network are written in C and executed on a PC under Red Hat Linux version 7.3 or 8.0.

Training of the ANN (step 4 of the algorithm) is generally executed only once, and the same trained neural network can be utilized to execute the method on any prokaryotic genome. Although if Applicants use organism specific training set, results might improve in some cases, but it would be marginal. This is because our method predicts gene on the basis of the number distribution of the alphanumeric sequence of an ORF. So the gene prediction is more dependent on the peptide library used rather than training set.

5. Deciphering genes using trained ANN

While creation of peptide library (step 1) and training of ANN (step 4) are considered as preparatory phases for executing the method of invention, step 2 and step 3 are mandatory for each genome sequence. After translating computationally a genome into all six reading frames and converting them into six alphanumeric sequences, deciphering genes using ANN is executed. This step can be further divided into following five sub-steps:

20

1. Breaking of all the six alphanumeric sequences into possible ORFs. (all possible fragments starting with 's' and ending with '*')

2. Calculate all the five parameters (total score, fraction of zeroes, mean, variance, and length of maximum continuous non zero stretch) for all possible ORFs (all the alphanumeric string sequences between 's' and '*').

3. Calculate the probability of the ORF corresponding to a given alphanumeric string as a protein coding gene, using the trained ANN.

4. Filter out the protein coding ORFs from the non coding ones by using a cutoff probability value.

5. Remove all the encapsulated protein coding regions (Shibuya,T. and Rigoutsos,I., 2002).

If two ORFs are predicted in distinct translation frames, such that one's span completely encapsulates other, it is a commonly believed that only one of them can be an actual gene. In this case the applicants report the ORF with a higher probability value as a gene. In case of same probability value Applicants take longer ORF as a gene.

The method of the invention predicts a probability value corresponding to a query ORF being a protein coding region. The training of ANN is done using a sigmoid learning function with = 1 (probability '1' for genes and '0' for non-genes); therefore most of the time this probability value lies either below 0.1 or above 0.9. Due to this any cutoff value lying between 0.1 and 0.9 generate very similar results. In our analysis Applicants use a default cutoff value of 0.5. It's important to note that the method does not require a trade-off between sensitivity and specificity because the choice of cut-off probability has no major consequences on the results.

D.D.

Motivation: The recent out break of Severe Acute Respiratory Syndrome caused by SARS coronavirus has necessitated in-depth molecular understanding of the virus to identify new drug targets. The availability of complete genome sequence of several strains of SARS virus provides the possibility of identification of protein coding genes and defining their functions. Computational approach to identify protein coding genes and their putative functions will help in designing experimental protocols.

21

Results: In this invention a novel analysis of SARS genome using gene prediction method GeneDecipher developed in our laboratory, has been presented. Each of the 18 newly sequenced SARS-CoV genomes has been analyzed using GeneDecipher. In addition to polyprotein 1ab*, polyprotein 1a and the four genes coding for major structural proteins spike(S), small envelope (E), membrane (M), and nucleocapsid (N), 6 to 8 additional proteins have been predicted depending upon the strain analyzed. Their lengths range between 61 and 274 amino acids. Our method also suggests that polyprotein spike (S), membrane (M), Nucleocapsid (N) are proteins of viral origin and others are of prokaryotic. Putative functions 01 all predicted protein coding genes have been suggested using conserved peptides present in their ORFs.

*GeneDecipher predicts polyprotein lab (265....21485) in two fragments (265...13413) and (13599. ..21485) because there is a stop codon at location 13413. These locations are given with respect to the NCBI refseq Genome sequence.

GeneDecipher originally developed for prokaryotic gene prediction, rjgeds only parameters and can therefore analyze smaller genomes too. Applicants have trained the Artificial Neural Network on *ecoli*-kl2 genome coding and non-coding regions (ORFs not reported as a gene). To predict protein coding genes using GeneDecipher on viral genomes no additional training is required. This is an obvious advantage of this method over other methods In addition it's very difficult to find negative training set (non-coding regions) for small genomes like coronavirus. Non-coding sequences for training are made by shuffling the coding sequences (Chen et al., 2003). The obviation of need to train specifically for the organism thus makes GeneDecipher suitable for such small genomes.

In continuation Applicants tried to assign function to the GeneDecipher predicted SARS-CoV genes using PLHOST, a tool for functional prediction developed at our laboratory PLHOST assigns function based upon the presence of invariant octa/hepta peptides across proteins from different species. In this invention Applicants present the results of our analysis on 18 SARS-CoV genomes.

Methods

*SARS-CoV genome sequence:*Sequences of the 18 SARS-CoV strains available in the GenBank database (http://www.ncbi.nlm.nih.gov/Entrez/genomes/viruses) were

downloaded   and   analyzed.
These include SARS-CoV Refseq (NC_004718.3),SARS-CoV TWC(AY32118), SIN2774(AY283798),SIN2748(AY283797)SIN267^(AY283796), SIN2677(AY283794), SIN25ti6(AY283794), Frankfurt1(A Y291315), BJ04(AY279354) BJ03(AY278490), BJ02(AY278487), GZ01(AY278848), CUHKW1(AY278554), TOR2(AY274119), TW1(AY291451), BJ01(AY278488), Urban(AY278741), HKU-39849(AY278491). Other information related to protein coding genes was retrieved from.http://www.ncbi.nlm.nih.gov/genomes/SARS/SAks.html

GeneDecipher: Protein coding gene prediction software (separate manuscript communicated)

Originally GeneDecipher was developed for prokaryotic gene prediction. To execute GeneDecipher on viral genomes Applicants prepared a heptapeptide library derived from proteins of 56 completely sequenced prokaryotic genomes and 1096 viral genomes.

Development of GeneDecipher is based upon the observation that difference between total number of theoretically possible peptides of a given length and that which are actually observed in nature, grows drastically as this length of peptide increases. Moreover, it is interesting to note that most of these peptides selected by nature are found only in coding regions and very rarely in theoretically translated non-coding regions. This observation has prompted us to exploit this exclusivity of natural selection of peptides that are present in protein coding sequences to differentiate between coding and non-coding regions.

Prediction of a given ORF as a coding region/gene is based upon the number of heptapeptides present and the distribution of these heptapeptide along the ORF. Our output corresponding to a given ORF it a probability value I probability of this QRF being a gene). The final cut-off probability is user dependent, but it is constant for a given genome in all six reading frames (default cut-off is 0.5).

Here it is worth noting that our method is independent of any other evidences, e.g ribosome binding site signals (in order to prove the strength of the hypothesis) such kinds of constraints are being used by various existing methods.

The method can be divided into Five major steps (Figure 1):

1. Generation of a peptide library.

2. Artificial translation of a given genome into 6 reading frames.

3. Conversion of each translated sequence into an integer coded sequence, (one corresponding to each reading frame).

4. Training of artificial neural network (ANN),

5 Deciphering genes using trained ANN.

PLHOST: Function Assignment Tool

Applicants used PLHOST {Peptide Library based Homology Search Tool) for the identification of invariant peptides which serve as functional signatures from completely sequenced genomes.

10 The algorithm generates organism specific libraries of octa/hepta peptides from all proteins of selected genomes. Redundant peptides are removed from each library. These peptide libraries are then compared with each other to note all octa/hepta peptides present invariantly across a specified minimum number of genomes, Overlapping octa/hepta peptides are back stitched to generate longer conserved peptides which occur in

15 functionally similar proteins, hence called functional signatures.

Results and Discussion:

A systematic sensitivity and specificity analysis of GeneDecipher has been done on 10 microbila genomes (Figure 2). Further analysis of GeneDecipher on viral genomes is presented here.

20 *Testing of GeneDecipher on viral genomes:*

To test our method on viral genomes Applicants first analyzed *Human Respiratory Syncytial Virus (HRSV)*, complete genome using GeneDecipher. Comparison of GeneDecipher results with state of the art method ZCURVE_CoV has been done (Table 1). ZCURVE_CoV is able to predict 8 annotate proteins out of 11 reported at NCBI

25 without any false positives, ZCURVE_CoV was unable to predict the following three genes: PID 9629200 (location 626....1000, non-structural protein 2 (NS2)); PID 9629205 (location 4690.....5589, attachment glycoprotein (G)); and PID 9629208 (location 8171 ....8443, matrix protein 2(M2)). GeneDecipher predicted 10 out of total 11 annotated proteins of HRSV without any false positives. The Gene missed by GeneDecipher was PID

30 9629208 (location 8171....8443, matrix protein 2) which was notably missed by ZCURVE_CoV too.

This successful prediction of protein coding regions in *HRSV* genome increases our confidence to predict protein coding regions on newly sequences SRAS_CoV genomes.

Analysis of SRAS-CoV using GeneDecipher:

Applicants analyzed all 18 strains of SARS-CoV using GeneDecipher. GeneDecipher predicts a total of 15 protein coding regions in SARS-CoV genomes including both the polyproteins 1a, 1ab (Sars2628 C-terminal end of Polyprotein1ab), and all four known structural proteins (M, N, S, and E) for each of the 18 strains. GeneDecipher also predicts 6 to 8 additional coding regions depending on the genome sequence of the strain used. The length of these additional coding regions varied between 61 and 274 amino acids.

GeneDecipher predicts 12 coding regions which are common to all 18 strains (Table 2), and one coding region (Sars63, sars6 at NCBI refseq genome) present in 5 strains. GeneDecipher predicts gene Sars90 in GZ01 strain, and Sars154 (Sars 3b at NCBI refseq genome) in BJ02 strain specifically.

These 12 common protein coding regions consist of the 6 basic proteins of SARS-CoV (2 polyproteins and the 4 structural proteins); Sars274 (Sars3a at NCBI refseq database), Sars122 (Sars7a at NCBI refseq database), Sars78 (already reported with start shifted as ORF14/Sars9c in TOR2strain); and three newly predicted (false positives with respect to current annotation at NCBI) protein coding regions Sars174, Sars68, and Sars61. The three newly predicted genes lie completely within polyprotein 1a genomic region. Although our method discards such genes in bacterial genomes, possibility of finding such genes in viral genomes has not been ruled out. As these genes are present in all 18 strains it is likely that they are protein coding genes.

Applicants predict three more coding regions Sars63, Sars154, and Sars90 apart from the 12 discussed above, Sars63 is identified in 5 strains and not identified in remaining 13 strains. This coding region is already reported in NCBI refseq (Sars6). Here Applicants can not comment much about the existence of Sars63 (Sars6 at NCBI refseq) because it is identified in 5 strains and not identified in rest 13. This is due to high density of non-synonymous mutations across strains in this region. Two coding regions Sars154 (sars3b at NCBI), and Sars90 (newly predicted in GZ01 strain) are identified in one strain. The locations of these three genes in different strains are provided in Table 3.

Since the peptide libraries are made from the genome sequences of various organisms, the evolutionary origin of a given protein can be traced. If the protein is rich in heptapeptides found occurring in viral genomes then that protein is considered to be of viral origin. Applicants found that 5 core proteins (two polyproteins and three structural proteins M, N, and S) are of viral origin. The remaining, including 3 new predictions are of prokaryotic origin. It is interesting to that from the same DNA region Applicants' are getting proteins in different frames which contain peptides from different origin. Here, how same DNA sequence can code for both bacterial and viral origin in intriguing. This might explain why these new protein coding genes were not detected in primary attempts based on homology to other known viral genome sequences.

Comparison with the existing system – ZCURVE_CoV:

Comparison of GeneDecipher, ZCURVE_CoV results with the known annotations for Urbani and TOR2 strains of SARS-CoV are presented in Tables 4a and 4b.

In general, GeneDecipher results are in good agreement with the known annotations. In case of Urbani strain GeneDecipher predicts all the known genes except Sars84(X5), Sars63(X3) and Sars154(X2). Sars84(X5) and Sars63(X3) are supported by ZCURVE_CoV whereas Sars154(X2) is missed by both the methods. GeneDecipher predicts four new genes in this strain which incidentally are not supported by ZCURVE_CoV. It is noticeable that out of these four genes Sars78 is already known for strain TOR2 as ORF14/Sars9c. This supports the likelihood of the gene being present in Urbani strain. However, ZCURVE_CoV predicts 2 new genes which are not supported by GeneDecipher either.

GeneDecipher predictions for TOR2 strain are identical with those for Urbani strain. In this strain GeneDecipher predicts 9 known genes but fails to predict 6 genes with known annotations. These 6 genes are: Sars154 (ORF4), Sars98 (ORF13), Sars63 (ORF7), Sars44 (ORF9), Sars39 (ORF10), and Sars84 (ORF11). Of these, Sars154 (ORF4) and Sars98 (ORF13) are also missed by ZCURVE_CoV. It is to be noted that both Sars44 (ORF9) and Sars39 (ORF10) are ORFs very small in length (44 and 39 amino acids respectively), and their presence too is not consistent across various SARS strains. Sars63 (ORF7) has been predicted by GeneDecipher in 5 other strains but not in the two strains considered here.

26

Mutation Analysis:

Analysis using multiple sequence alignment (ClustalW) for 3 newly predicted protein coding genes Sars174, Sars68 and Sars61 across all 18 strains shows:

1. Sars68 has one point mutation at location80 GAT->GGT (D->G) Sin2677 strain.
2. 2. Sars 174 has two synonymous point utations at location 204 CGA->CGC in GZ01 strain and at location 447 CTG->CTT in BJ04 strain.
3. 3. Sars 61 has one point mutation at location 119 CTG->CAG (L->Q) in GZ01 strain.

These three newly predicted genes are present in all 18 strains without significant mutations and has no significant hits with BLASTP in non-redundant database. This indicates that these three proteins might have crucial biological functions specific to SARS-CoV. Therefore these coding sequences might serve as candidate drug targets against SARS.

Function Assignment

In total Applicants predict 15 coding regions in SARS-CoV out of which functions of the four structural proteins (M, N, S and E) have already been assigned. Although the polyprotein lab has been assigned only replicase activity, our analysis implies that the replicase activity is associated with Sars2628 (C terminal of ORF 1ab) fragment. The complete 1ab polyprotein contains 6 functional signatures of which polyprotein 1a contains signatures associated with metabolic enzymes [Table 5a]. Functions were assigned to the polyproteins on the basis of peptides (length 7 or more amino acids) occurring in proteins having similar functions in at least 5 different organisms. Other predicted genes/protein coding regions contain peptides which occur in fewer genomes. Based on these peptides Applicants suggest functions (Table 5b). The biological relevance of these finding remains to be explored.

Conclusion:

In this application applicants have predicted 4 new genes including Sars78 (already known in TOR2 strain) in SARS-CoV. Our analysis also corroborates the finding of ZCURVE_CoV (Can et al, 2003) that ORF Sars154 (listed in Refseq as Sars3b) is unlikely to be a coding region. Applicants have also assigned functions to the two polyproteins 1ab and 1a. In addition to replication associated function of C-terminal of 1ab polyprotein, our analysis implies that the polyprotein 1a may be associated with metabolic enzyme like

27

functions. In all, six peptide signatures are present in polyprotein 1ab. Applicants have suggested putative function for other 9 proteins including ones newly predicted by GeneDecipher.

Table1 Comparison of GeneDecipher results with ZCURVE_CoV results on HRSV genome, with respect to annotated genes

Annotated genes			ZCURVE CoV			GeneDecipher		
Start	End	Length	Start	End	Length	Start	End	Length
99	518	139	99	518	139	99	518	139
626	1000	124	—	—	—	626	1000	124
1140	2315	391	1140	2315	391	1140	2315	391
2348	3073	241	2348	3073	241	2348	3073	241
3263	4033	256	3158	4033	291	3158	4033	291
4303	4500	65	4303	4500	65	4303	4500	65
4690	5589	299	—	—	—	4690	5589	299
5666	7390	574	5666	7390	574	5621	7390	589
7618	8205	195	7618	8205	195	7618	8205	195
8171	8443	90	—	—	—	—	—	—
8509	15009	2166	8443	15009	2188	8443	15009	2188

Table 2: Protein coding genes predicted by GeneDecipher in SARS-CoV Refseq common to all 18 strains.

S.No.	Start	Stop	Frame	Length		Feature
				bp	aa	
1	265	13413	1+	13149	4382	Sars 1a polyprotein
2	701	1225	2+	525	174	Sars174(new prediction)
3	1397	1603	2+	207	68	Sars68(new prediction)
4	8828	9013	2+	186	61	Sars61(new prediction)
5	13599	21485	3+	7887	2628	Sars2628(C-terminal end of polyprotein 1ab)
6	21492	25259	3+	3768	1255	Spike (S) protein
7	25268	26092	2+	825	274	Sars274(Sars 3a)
8	26117	26347	2+	231	76	Sars76(Sars4)
9	26398	27063	1+	666	221	Sars221(Sars5)
10	27273	27641	3+	369	122	Sars122(Sars7a)
11	28120	29388	1+	1269	422	Sars422(Sars9a)
12	28559	28795	2+	237	78	Sars78(Identical to ORF 14/Sars9c in TOR2 with shifted start)

Table3: Identification of Sars90, Sars63, Sars154 as protein coding genes by GeneDecipher in various strains of SARS-CoV

S.No.	Strain name	Sars90 (new prediction)	Sars63(Sars6 at NCBI)	Sars154(Sars 3b at NCBI)
1	SARS_2748	--	--	--
2	SARS_bj01	--	27055..27246	--
3	SARS_bj02	--	27074..27265	25689..26153
4	SARS_bj03	--	27070..27261	--
5	SARS_bj04	--	27058..27249	--
6	SARS_frankft 1	--	----	--
7	SARS_urbani	--	--	--
8	SARS_gz01	24492..24764	27058..27249	--
9	SARS_sin2500	--	--	--
10	SARS_sin2677	--	--	--
11	SARS_sin2679	--	--	--
12	SARS_sin2774	--	--	--
13	SARS_chuk	--	--	--
14	SARS_tw1	--	--	--
15	SARS_twc	--	--	--
16	SARS_hku39849	--	--	--
17	SARS_refseq	--	--	--
18	SARS_TOR2	--	--	--

Table 4(a). Comparison of GeneDecipher results with ZCURVE_CoV results on SARS-CoV genome Urbani strain, with respect to annotated genes

Annotated genes			ZCURVE_CoV			GeneDecipher			Features
Start	End	Length	Start	End	Length	Start	End	Length	
265	13398	4377	265	13398	4377	265	13413	4382	ORF 1a
--	--	--	--	--	--	701	1225	174	Sars174(New prediction by GeneDecipher)
--	--	--	--	--	--	1397	1603	68	Sars68(New prediction by GeneDecipher)
--	--	--	--	--	--	8828	9013	61	Sars61(New prediction by GeneDecipher)
13398	21485	2695	13398	21485	r~2691	13599	21485	2628	ORF1b
21492	25259	1255	21492	25259	1255	21492	25259	1255	S protein
25268	26092	274	25268	26092	274	25268	26092	274	Sars274(XI)
25689	26153	154	--	--	--	--	--	--	Sars154(X2)
26117	26347	76	26117	26347	76	26117	26347	76	E protein
26398	27063	221	26398	27063	221	26389	27063	224	M protein
27074	27265	63	27074	27265	63	--	--	--	Sars63(X3)
27273	27641	122	27273	27641	122	27273	27641	122	Sars122(X4)
~	—	—	27638	27772	H^	-	—	—	Sars44
—	—	~	27779	27898	39	--	--	--	Sars39
27864	28118	84	27864	~28rnr	84	~	--	—	Sars84(X5)
|28120	~2938lTl	422	28120	29388	422	28120	29388	422	N protein
—	—	—	--	—	,,	28559	28795	78	Sars78(Identical to ORF 14/Sars9c in TOR2 with shifted start)

Table 4(b). Comparison of GeneDecipher results with ZCURVE_CoV results on SARS-CoV genome TOR2 strain, with respect to annotated genes

Annotated genes			ZCURVE_CoV predicted genes			GeneDecipher predicted genes			Features
Start	End	Length	Start	End	Length	Start	End	Length	
265	13398	4377	265	13398	4377	265	13413	4382	ORF la
--	--	--	--	--	--	701	1225	174	Sars174(New prediction by GeneDecipher)
--	--	--	--	--	--	1397	1603	68	Sars68(New prediction by GeneDecipher)
--	--	--	--	--	--	S828	9013	61	Sars61(New prediction by GeneDecipher)
13398	21485	2695	13398	21485	2695	13599	21485	2628	ORF lb
21492	25259	1255	21492	25259	1255	21492	25259	1255	S protein
25268	26092	274	25268	26092	274	25268	26092	274	ORF3(Sars274)
25689	26153	154	--	--	--	--	--	--	ORF4(Sars154)
26117	26347	76	26117	26347	76	26117	26347	76	E protein
26398	27063	221	26398	27063	221	26389	27063	224	M protein
27074	27265	63	27074	27265	63	--	--	--	Sars63(ORF7)
27273	27641	122	27273	27641	122	27273	27641	122	Sars122(ORF8)
27633	27772	44	27638	27772	44	--	--	--	Sars44(ORF9)
27779	27898	39	27779	27898	39	--	--	--	Sars39(ORF10)
27864	28118	84	27864	28118	84	--	--	--	Sars84(ORFll)
28120	29388	422	28120	29388	422	28120	29388	422	N protein
28130	28426	98	--	--	--	--	--	--	ORF13
28583	28795	70	--	--	--	28559	28795	78	Sars78(Identical toORF 14/Sars9c in TOR2 with shifted start)

Table 5(a): Functional assignment of polyproteins at SARS (Urbani) Genome using PLHOST

S.No.	NCBI annotation	Conserved peptide signature	Function assigned
1	Sars 1ab (poly protein 1ab)		Phosphoglycerate kinase
		RSETLLPL	Sulfite reductase (NADPH), Flavoprotein beta subunit
		LDKLKSLL	Probable acyl-CoA thiolase
		ATWIGTS	cell division protein ftsZ
		NVATTRAK	DNA-binding protein, probably DNA helicase
		LQGPPGTGK	DNA helicase related protein
2	Sars 1a poly protein 1a	RIRASLPT	Phosphoglycerate kinase
		RSETLLPL	Sulfite reductase (NADPH), Flavoprotein beta subunit
		LDKLKSLL	Probable acyl-CoA thiolase
3	Sars 2628 (C terminal of Sars 1ab)	ATWIGTS	cell division protein ftsZ
		NVAITRAK	DNA-binding protein, probably DNA helicase
		LQGPPGTGK	DNA helicase related protein

Tables (b): Suggested functions for some of the non-structural genes in SARS-CoV using PLHOST

S.No.	Gene	Peptide Signature	Suggested function
1	Sars174(new prediction)	TLSKGNAQ	ABC transporter ATP binding protein *[Lactococcus lactis subsp. lactis]*
		VAQMGTLL	Cytochrome c oxidase folding protein *[Synechocystis sp. PCC 6803]*
2	Sars68(new prediction)	LVLVLILA	*putative major facilitator superfamily protein [Schizosaccharomyces pombe]*
		TQTLKLDS	*serine/threonine kinase 2; Serine/threonine protein hnase-2 [Homo sapiens]*
3	Sars90(new prediction only in GZ01 strain)	GLLHRGT	NADH Dehydrogenase I Chain
4	Sars61(new prediction)	LLPLLAFL	Putative protein (Conserved across 1 organisms)
5	Sars274(Sars3a)	LLLFVTIY	Polyamine transport protein; Tpolp *[Saccharomyces cerevisiae]*
6	Sarel54(Sars3b)	QTLVLKML	K550.3.p *[Caenorhabditis elegans]*
7	Sars63(Sars6)	DDEELMEL	Elongation factor Tu *[Lactococcus lactis subsp. lactis]*
8	Sars122(Sars7a)	LIVAALVF	Putative transport transmembrane protein *[Sinorhizobium melilotij]*
		RARSVSPK	Src homology domain 3 *[Caenorhabditis elegans]*
9*	Sars78(Sars9c)	QLLAAVG	Gamma-glutamate kinase (Conserved across 8 organisms)

*: No conserved octapeptide was found. However, function has been assigned on the basis of the only highly conserved heptapeptide.

Table 6: GeneDecipher Prediction on *SARS-CoV 2748* strain

S.No.	start	end	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307
11	28099	29367	1+	422	0.927307
12	28538	28774	2+	78	0.927307

Table 7: GeneDecipher Prediction on *SARS-CoV BJ01* strain

S.No.	start	End	frame	length	Probability
1	246	13394	3+	4382	0.927307
2	682	1206	1+	174	0.927307
3	1378	1584	1+	68	0.927307
4	8809	8994	1+	61	0.927307
5	13580	21466	2+	2628	0.927307
6	21473	25240	2+	1255	0.927307
7	25249	26073	1+	274	0.927307
8	26098	26328	1+	76	0.925291
9	26370	27044	3+	224	0.927307
10	27254	27622	2+	122	0.927307
11	28101	29369	3+	422	0.927307
12	28540	28776	1+	78	0.927307

Table 8: GeneDecipher Prediction on *SARS-CoV BJ02* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	25689	26153	3+	154	0.927268
9	26117	26347	2+	76	0.925291

10	26389	27063	1+	224	0.927307
11	27273	27641	3+	122	0.927307
12	28120	29388	1+	422	0.927307
13	28559	28795	2+	78	0.927307

Table 9: GeneDecipher Prediction on *SARS-CoV BJ03* strain

S.No.	start	end	frame	length	Probability
1	261	13409	3+	4382	0.927307
2	697	1221	1+	174	0.927307
3	1393	1599	1+	68	0.927307
4	8824	9009	1+	61	0.927307
5	13595	21481	2+	2628	0.927307
6	21488	25255	2+	1255	0.927307
7	25264	26088	1+	274	0.927307
8	26113	26343	1+	76	0.925291
9	26385	27059	3+	224	0.927307
10	27269	27637	2+	122	0.927307
11	28116	29384	3+	422	0.927307
12	28555	28791	1+	78	0.927307

Table 10: GeneDecipher Prediction on *SARS-CoV BJ04* strain

S.No.	start	end	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307
11	28104	29372	3+	422	0.927307
12	28543	28779	1+	78	0.927307

Table 11: GeneDecipher Prediction on *SARS-CoV CHUK* strain

S.No.	start	end	frame	length	Probability
1	250	13398	1+	4382	0.927307
2	686	1210	2+	174	0.927307
3	1382	1588	2+	68	0.927307
4	8813	8998	2+	61	0.927307
5	13584	21470	3+	2628	0.927307
6	21477	25244	3+	1255	0.927307
7	25253	26077	2+	274	0.927307
8	26102	26332	3+	76	0.925291
9	26374	27048	1+	224	0.927307
10	27258	27626	3+	122	0.927307

11	28105	29373	1+	422	0.927307
12	28544	28780	2+	78	0.927307

Table 12: GeneDecipher Prediction on *SARS-CoV Frankfurt1* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

Table 13: GeneDecipher Prediction on *SARS-CoV GZ01* strain

S.No.	start	end	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	24492	24764	3+	90	0.927307
8	25252	26076	1+	274	0.927307
9	26101	26331	1+	76	0.927307
10	26373	27047	3+	224	0.927307
11	27058	27249	1+	63	0.927307
12	27257	27625	2+	122	0.927307
13	28133	29401	2+	422	0.927307
14	28572	28808	3+	78	0.927307

Table 14: GeneDecipher Prediction on *SARS-CoV HKU39849* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307

10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

Table 15: GeneDecipher Prediction on *SARS-CoV Refseq* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

Table 16: GeneDecipher Prediction on *SARS-CoV SIN2500* strain

S.No.	start	End	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307
11	28104	29372	3+	422	0.927307
12	28543	28779	1+	78	0.927307

Table 17: GeneDecipher Prediction on *SARS-CoV SIN2677* strain

S.No.	start	End	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307

| 11 | 28098 | 29366 | 3+ | 422 | 0.927307 |
| 12 | 28537 | 28773 | 1+ | 78 | 0.927307 |

Table 18: GeneDecipher Prediction on *SARS-CoV SIN2679* strain

S.No.	start	end	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307
11	28104	29372	3+	422	0.927307
12	28543	28779	1+	78	0.927307

5   **Table 19: GeneDecipher Prediction on *SARS-CoV SIN2774* strain**

S.No.	start	end	frame	length	Probability
1	249	13397	3+	4382	0.927307
2	685	1209	1+	174	0.927307
3	1381	1587	1+	68	0.927307
4	8812	8997	1+	61	0.927307
5	13583	21469	2+	2628	0.927307
6	21476	25243	2+	1255	0.927307
7	25252	26076	1+	274	0.927307
8	26101	26331	1+	76	0.925291
9	26373	27047	3+	224	0.927307
10	27257	27625	2+	122	0.927307
11	28104	29372	3+	422	0.927307
12	28543	28779	1+	78	0.927307

Table 20: GeneDecipher Prediction on *SARS-CoV TOR2* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

Table 21: GeneDecipher Prediction on *SARS-CoV TW1* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

Table 22: GeneDecipher Prediction on *SARS-CoV TWC* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28118	29386	2+	422	0.927307
12	28557	28793	3+	78	0.927307

Table 23: GeneDecipher Prediction on *SARS-CoV Urbani* strain

S.No.	start	end	frame	length	Probability
1	265	13413	1+	4382	0.927307
2	701	1225	2+	174	0.927307
3	1397	1603	2+	68	0.927307
4	8828	9013	2+	61	0.927307
5	13599	21485	3+	2628	0.927307
6	21492	25259	3+	1255	0.927307
7	25268	26092	2+	274	0.927307
8	26117	26347	2+	76	0.925291
9	26389	27063	1+	224	0.927307
10	27273	27641	3+	122	0.927307
11	28120	29388	1+	422	0.927307
12	28559	28795	2+	78	0.927307

The particulars of the organisms such as their name, strain, accession number and other details are given below.

S.No.	Genome	Strain	Accession Number	Total Base Sequences	Date of Completion
1	*H.Influenzae*	Rd	NC_000907	1830138	Sep30,1996

Fleischmann,R.D. et.al Science 269 (5223), 496-512 (1995)

| 2 | *M.Genitalium* | -- | NC_000908 | 580074 | Jan8,2001 |

Fraser,C.M., et.al Science 270 (5235), 397-403 (1995

| 3 | *E.coli* | K-12 | NC_000913 | 4639221 | Oct 15, 2001. |

Blattner,F.R. et. al Science 277 (5331), 1453-1474 (1997)

| 4 | *B. Subtilis* | 168 | NC_000964 | 4214814 | Nov 20,1997 |

Kunst,F. et.al Nature 390 (6657), 249-256 (1997)

| 5 | *A.Fulgidis* | DSM 4304 | NC_000917 | 2178400 | Dec.17,1997 |

Klenk,H.P.et.al Nature 390 (6658), 364-370 (1997)

| 6 | *M.Tuberculosis* | H37RV | NC_000962 | 4411529 | Sep.7,2001 |

Cole,S.T. et.al Nature 393 (6685), 537-544 (1998)

| 7 | *T.Pallidum* | -- | NC_000919 | 1138011 | Sep 7, 2001 |

Fraser,C.M.,et.al Science 281 (5375), 375-388 (1998)

| 8 | *T.Maritima* | -- | NC_000853 | 1860725 | Sep 10, 2001. |

Nelson,K.E. et.al Nature 399 (6734), 323-329 (1999)

| 9 | *Synecho cystis* | PCC6803 | NC_000911 | 3573470 | Oct 30,1996 |

41

Kaneko,T. et.al DNA Res. 3(3), 109-136 (1996)

10   *H.Pylori*   26695   NC_000915   1667867
      Sep7,2001

Tomb,J.-F. et.al Nature 388 (6642), 539-547 (1997)

5   In another embodiment of the present invention, wherein further, certain organisms were studied in detail using the method of the instant Application. The gene coding regions of the same were identified and also, their putative functions. The same is reflected in the following 165 sequences. They are placed in a sequential order starting from SEQ ID No. 9 to SEQ ID No 173. The details are as given below.

10

Tabular Format

Table No. 24: Organism name: Haemophilus influenzae

S.No	GDC ID	Start	End	Length	Frame	Putative function
1	GDC_HINF_5641	5641	6273	210	+	Formate dehydrogenase major subunit
2	GDC_HINF_6322	6322	8748	808	+	Formate dehydrogenase major subunit
3	GDC_HINF_124181	124181	124378	65	+	Cell wall-associated hydrolase
4	GDC_HINF_170553	170553	170732	59	-	dicarboxylate transport protein homolog HI0153
5	GDC_HINF_231874	231874	232173	99	+	type I restriction system adenine methylase
6	GDC_HINF_232170	232170	232991	273	+	type I restriction system adenine methylase
7	GDC_HINF_232813	232813	233139	108	+	type I restriction system adenine methylase
8	GDC_HINF_233190	233190	233393	67	+	Type I restriction enzyme EcoprrI M protein
9	GDC_HINF_235441	235441	235932	163	+	prrD protein homolog

10	GDC_HINF_235913	235913	238519	868	+	Type I restriction enzyme EcoR124II R protein
11	GDC_HINF_240336	240336	241379	347	-	Aerobic respiration control sensor protein
12	GDC_HINF_243018	243018	243215	65	+	Cell wall-associated hydrolase
13	GDC_HINF_274892	274892	276853	653	-	Adhesion and penetration protein precursor
14	GDC_HINF_276992	276992	279121	709	-	Adhesion and penetration protein precursor
15	GDC_HINF_370413	370413	370808	131	+	NapA
16	GDC_HINF_370747	370747	372912	721	+	NapA
17	GDC_HINF_628407	628407	628604	65	-	Cell wall-associated hydrolase
18	GDC_HINF_654365	654365	655015	216	-	Probable D-methionine transport system permease
19	GDC_HINF_661444	661444	661641	65	-	Cell wall-associated hydrolase
20	GDC_HINF_737160	737160	737297	45	+	glycerophosphodiester phosphodiesterase
21	GDC_HINF_775792	775792	775989	65	-	Cell wall-associated hydrolase
22	GDC_HINF_848166	848166	848678	170	-	ribosomal protein
23	GDC_HINF_928073	928073	929080	335	+	Peptidase B (Aminopeptidase B)
24	GDC_HINF_929037	929037	929402	121	+	Peptidase B (Aminopeptidase B)
25	GDC_HINF_1018846	1018846	1021371	841	-	Isoleucyl-tRNA synthetase
26	GDC_HINF_1021582	1021582	1021683	33	-	Isoleucyl-tRNA synthetase

27	GDC_HINF_1082407	1082407	1082514	35	-	protein V6, truncated - Haemophilus influenzae
28	GDC_HINF_1144501	1144501	1145004	167	-	PnuC transporter
29	GDC_HINF_1279189	1279189	1279935	248	-	Peptide chain release factor 2 (RF-2)
30	GDC_HINF_1347200	1347200	1347445	81	+	putative ABC transport protein
31	GDC_HINF_1347942	1347942	1348478	178	+	putative iron compound ABC transporter
32	GDC_HINF_1476415	1476415	1476615	66	-	PstB
33	GDC_HINF_1476557	1476557	1477183	208	-	PstB
34	GDC_HINF_1505851	1505851	1506048	65	-	terminase large subunit
35	GDC_HINF_1524561	1524561	1525421	286	-	ThiI
36	GDC_HINF_1568974	1568974	1569300	108	+	DNA-binding protein rdgB homolog
37	GDC_HINF_1586944	1586944	1587765	273	+	putative tail protein
38	GDC_HINF_1594339	1594339	1594854	171	-	NifC
39	GDC_HINF_1634710	1634710	1636722	670	+	Probable hemoglobin and hemoglobin-haptoglobin
40	GDC_HINF_1638626	1638626	1639372	248	-	Putative integrase/recombinase HI1572
41	GDC_HINF_1639409	1639409	1639726	105	-	Putative integrase/recombinase HI1572
42	GDC_HINF_1660491	1660491	1662080	529	-	Cell division protein ftsK homolog
43	GDC_HINF_1807963	1807963	1808859	298	-	adhesin homolog HI1732
44	GDC_HINF_1817220	1817220	1817417	65	+	Cell wall-associated hydrolase

Table No 25: Organism Name: Helicobacter pylori

S.No.	GDC ID	Start	End	Length	Frame	Putative function
1	GDC_HPYL_51094	51094	51432	112	-	putative HP0052-like protein
2	GDC_HPYL_155367	155367	156164	265	-	2-oxoglutarate/malate translocator
3	GDC_HPYL_447632	447632	447850	72	-	Cell wall-associated hydrolase
4	GDC_HPYL_506250	506250	507134	294	+	site-specific DNA-methyltransferase
5	GDC_HPYL_583607	583607	583876	89	+	probable DNA helicase
6	GDC_HPYL_583883	583883	584437	184	+	probable DNA helicase
7	GDC_HPYL_665045	665045	665695	216	+	putative lipopolysaccharide biosynthesis protein
8	GDC_HPYL_953783	953783	954664	293	-	acetate kinase
9	GDC_HPYL_954679	954679	954900	73	-	phosphate acetyltransferase
10	GDC_HPYL_954846	954846	955217	123	-	PHOSPHOTRANSACETYLASE
11	GDC_HPYL_955261	955261	955557	98	-	phosphate acetyltransferase
12	GDC_HPYL_1068602	1068602	1069459	285	-	IS606 TRANSPOSASE
13	GDC_HPYL_1069456	1069456	1069929	157	-	transposase-like protein, PS3IS
14	GDC_HPYL_1376803	1376803	1377126	107	+	ribosomal protein
15	GDC_HPYL_1474291	1474291	1474509	72	+	Cell wall-associated hydrolase
16	GDC_HPYL_1600102	1600102	1600689	195	-	TYPE III DNA MODIFICATION ENZYME

Table No. 26: Organism Name: Mycobacterium tuberculosis

S.No	GDC ID	Start	End	Length	Frame	Putative function
1	GDC_MTUB_26830	26830	27534	234	-	putative protoporphyrinogen oxidase
2	GDC_MTUB_36276	36276	36785	169	-	fibronectin-attachment protein FAP-P
3	GDC_MTUB_76032	76032	76595	187	+	retinoblastoma inhibiting gene 1
4	GDC_MTUB_80423	80423	81214	263	-	mucin 5
5	GDC_MTUB_167239	167239	168084	281	+	putative secreted peptidase
6	GDC_MTUB_214625	214625	215116	163	-	glycoprotein gp2
7	GDC_MTUB_424142	424142	424657	171	-	PPE FAMILY PROTEIN
8	GDC_MTUB_459316	459316	461076	586	+	63 kDa protein
9	GDC_MTUB_549643	549643	550758	371	-	carR
10	GDC_MTUB_566823	566823	567284	153	+	MAPK-interacting

						and spindle-stabilizing protein
11	GDC_MTUB_591109	591109	591345	78	+	excisionase, putative
12	GDC_MTUB_663028	663028	663426	132	+	PROBABLE RIBONUCLEOSIDE -DIPHOSPHATE REDUCTASE
13	GDC_MTUB_688806	688806	689060	84	+	MCE-FAMILY PROTEIN MCE2B
14	GDC_MTUB_701762	701762	702643	293	-	U1764ad
15	GDC_MTUB_731710	731710	731877	55	+	ribosomal protein L33
16	GDC_MTUB_772761	772761	773402	213	-	ENSANGP00000004 917
17	GDC_MTUB_868821	868821	869216	131	-	cold-shock induced protein of the Srp1p/Tip1p
18	GDC_MTUB_890358	890358	891254	298	-	orf2
19	GDC_MTUB_904043	904043	904840	265	+	aminoimidazole ribotide synthetase
20	GDC_MTUB_1045383	1045383	1046129	248	+	u650i
21	GDC_MTUB_1068100	1068100	1068726	208	-	anchorage subunit of a-agglutinin; Aga1p
22	GDC_MTUB_1115707	1115707	1116369	220	-	mucin 7 precursor, salivary
23	GDC_MTUB_1124996	1124996	1125712	238	-	putative oxidoreductase
24	GDC_MTUB_1138949	1138949	1139665	238	-	platelet binding protein GspB
25	GDC_MTUB_1170285	1170285	1170749	154	-	MC8
26	GDC_MTUB_1176592	1176592	1176858	88	+	gp85
27	GDC_MTUB_1202653	1202653	1203198	181	-	s19 chorion protein
28	GDC_MTUB_1231843	1231843	1232460	205	+	carboxylesterase
29	GDC_MTUB_1241031	1241031	1241468	145	-	PE
30	GDC_MTUB_1252888	1252888	1253748	286	-	ppg3
31	GDC_MTUB_1264312	1264312	1264554	80	+	ketoacyl-CoA thiolase-related protein
32	GDC_MTUB_1286282	1286282	1286587	101	-	pterin-4-alpha-carbinolamine dehydratase
33	GDC_MTUB_1301742	1301742	1302053	103	-	similar to ORF starts at 87, first start codon
34	GDC_MTUB_1351907	1351907	1352614	235	-	ppg3
35	GDC_MTUB_1476279	1476279	1476647	122	-	Cell wall-associated hydrolase
36	GDC_MTUB_1485311	1485311	1486399	362	-	4-hydroxyphenylpyruv ate dioxygenase C terminal
37	GDC_MTUB_1486309	1486309	1487727	472	-	cell wall surface anchor family protein
38	GDC_MTUB_1515112	1515112	1515846	244	-	putative ABC

					+	transporter ATP binding protein
39	GDC_MTUB_1515464	1515464	1516198	244	-	extracellular protein, gamma-D-glutamate-meso-d...
40	GDC_MTUB_1596569	1596569	1596892	107	-	putative translation initiation factor IF-2
41	GDC_MTUB_1600905	1600905	1601861	318	-	carboxylesterase family protein
42	GDC_MTUB_1616064	1616064	1616951	295	-	PUTATIVE TRANSCRIPTION REGULATOR PROTEIN
43	GDC_MTUB_1672449	1672449	1673216	255	+	MAV278
44	GDC_MTUB_1673708	1673708	1675000	430	-	MAV301
45	GDC_MTUB_1699549	1699549	1700226	225	+	gmdA
46	GDC_MTUB_1742061	1742061	1742858	265	-	ENSANGP00000020758
47	GDC_MTUB_1782153	1782153	1782932	259	+	GLP_26_54603_52153
48	GDC_MTUB_2060659	2060659	2061114	151	+	nuclear factor of kappa light polypeptide gene
49	GDC_MTUB_2093062	2093062	2093994	310	-	PROBABLE 6-PHOSPHOGLUCONATE DEHYDROGENASE GND1
50	GDC_MTUB_2105797	2105797	2106912	371	+	ATP-binding subunit of ABC-transport system
51	GDC_MTUB_2133554	2133554	2134069	171	-	KIAA0324 protein
52	GDC_MTUB_2183418	2183418	2184026	202	-	putative transport protein
53	GDC_MTUB_2192571	2192571	2193488	305	-	putative oxidoreductase
54	GDC_MTUB_2234641	2234641	2234889	82	-	DNA-binding protein, CopG family
55	GDC_MTUB_2320829	2320829	2321062	77	+	DNA-binding protein, CopG family
56	GDC_MTUB_2321250	2321250	2322509	419	-	cell wall surface anchor family protein
57	GDC_MTUB_2487508	2487508	2488524	338	-	ORF1
58	GDC_MTUB_2567990	2567990	2568457	155	+	B1158F07.3
59	GDC_MTUB_2577106	2577106	2577699	197	+	POSSIBLE CONSERVED MEMBRANE PROTEIN
60	GDC_MTUB_2577486	2577486	2577920	144	+	POSSIBLE CONSERVED MEMBRANE PROTEIN
61	GDC_MTUB_2690012	2690012	2690509	165	+	PROBABLE CONSERVED

						INTEGRAL MEMBRANE PROTEIN
62	GDC_MTUB_2698040	2698040	2698243	67	-	POSSIBLE CONSERVED MEMBRANE PROTEIN
63	GDC_MTUB_2712275	2712275	2714008	577	+	MLCL536.10 protein
64	GDC_MTUB_2725593	2725593	2725859	88	-	PROBABLE HYDROGEN PEROXIDE-INDUCIBLE GENES
65	GDC_MTUB_2733212	2733212	2734420	402	-	lycoprotein gp2
66	GDC_MTUB_2828257	2828257	2828937	226	+	MC8
67	GDC_MTUB_2895354	2895354	2897222	622	+	antigen T5
68	GDC_MTUB_2983047	2983047	2984033	328	-	MC8
69	GDC_MTUB_3005316	3005316	3005696	126	-	ABC transporter, ATP-binding protein
70	GDC_MTUB_3048559	3048559	3049095	178	-	recX protein
71	GDC_MTUB_3065095	3065095	3066549	484	+	ppg3
72	GDC_MTUB_3100192	3100192	3100452	86	-	IS1537, transposase
73	GDC_MTUB_3129118	3129118	3129594	158	-	KIAA1139 protein
74	GDC_MTUB_3237815	3237815	3238096	93	-	acylphosphatase
75	GDC_MTUB_3283182	3283182	3283718	178	-	Putative mycocerosyl transferase in MAS 5'r...
76	GDC_MTUB_3289702	3289702	3290232	176	+	POSSIBLE TRANSPOSASE
77	GDC_MTUB_3319076	3319076	3319546	156	-	U0002d
78	GDC_MTUB_3339006	3339006	3339851	281	-	membrane glycoprotein
79	GDC_MTUB_3356995	3356995	3357831	278	-	sensor histidine kinase
80	GDC_MTUB_3381198	3381198	3381755	185	+	MC8
81	GDC_MTUB_3388071	3388071	3389003	310	+	cellulosomal scaffoldin anchoring protein C
82	GDC_MTUB_3482312	3482312	3482770	152	-	MC8
83	GDC_MTUB_3581973	3581973	3582620	215	+	similar to mucin, submaxillary - pig
84	GDC_MTUB_3711717	3711717	3712613	298	-	orf2
85	GDC_MTUB_3716987	3716987	3718534	515	-	similar to profilaggrin – human (fragments)
86	GDC_MTUB_3754581	3754581	3755711	376	-	putative transposase
87	GDC_MTUB_3794808	3794808	3795026	72	-	deoxyxylulose-5-phosphate synthase
88	GDC_MTUB_3796793	3796793	3797512	239	+	membrane glycoprotein [imported] - equine herpesvirus
89	GDC_MTUB_3879013	3879013	3879534	173	-	ribosomal protein

						S11
90	GDC_MTUB_3921024	3921024	3921665	213	-	3-oxoacyl-(acyl-carrier-protein) reductase
91	GDC_MTUB_3974481	3974481	3975056	191	+	mucin.10
92	GDC_MTUB_3994808	3994808	3995446	212	+	MAV278
93	GDC_MTUB_3998938	3998938	3999642	234	-	protease inhibitor/seed storage/lipid transfer
94	GDC_MTUB_4021183	4021183	4021425	80	-	PUTATIVE TRNA/RRNA METHYLTRANSFE RASE
95	GDC_MTUB_4045946	4045946	4046290	114	-	chalcone/stilbene synthase family protein
96	GDC_MTUB_4053033	4053033	4053635	200	+	putative protein (2G313)
97	GDC_MTUB_4140236	4140236	4140460	74	-	DNA-binding protein, CopG family
98	GDC_MTUB_4169350	4169350	4169706	118	+	PROBABLE CUTINASE PRECURSOR CUT5
99	GDC_MTUB_4170798	4170798	4171211	137	+	PUTATIVE OXIDOREDUCTAS E
100	GDC_MTUB_4252190	4252190	4252921	243	+	Salivary gland secretion 1 CG3047-PA
101	GDC_MTUB_4260620	4260620	4261213	197	+	SPAPB15E9.01c
102	GDC_MTUB_4302166	4302166	4302858	230	+	u1764ad
103	GDC_MTUB_4317863	4317863	4318309	148	+	POSSIBLE TRANSPOSASE [SECOND PART]
104	GDC_MTUB_4341852	4341852	4342388	178	-	GLP_49_64409_654 43
105	GDC_MTUB_4391527	4391527	4391988	153	-	AT9S

In yet another embodiment of the present invention conserved peptide motifs as identified comprising using the instant methodology. They are present in a sequential order as amino acid sequences of SEQ ID Nos. 174 to 240.

1. AAQSIGEPGTQLT

2. AGDGTTTAT

3. AGRHGNKG

4. AHIDAGKTTT

5. CPIETPEG

6. DEPSIGLH

7. DEPTSALD

8. DEPTTALDVT

9. DHAGIATQ

10. DHPHGGGEG

11. DLGGGTFD

12. DVLDTWFSS

13. ERERGITI

14. ERGITITSAAT

15. ESRRIDNQLRGR

16. FSGGQRQR

17. GEPGVGKTA

18. GFDYLRDN

19. GHNLQEHS

20. GIDLGTTNS

21. GINLLREGLD

22. GIVGLPNVGKS

23. GKSSLLNA

24. GLTGRKIIVDTYG

25. GPPGTGKTLLA

26. GPPGVGKT

27. GSGKTTLL

28. GTRIFGPV

29. IDTPGHVDFT

30. IIAHIDHGKSTL

31. INGFGRIGR

32. IREGGRTVG

33. IVGESGSGKS

34. KFSTYATWWI

35. KMSKSKGN

36. KMSKSLGN

37. KNMITGAAQMDGAILVV

38. KPNSALRK

39. LFGGAGVGKTV

40. LGPSGCGK

41. LHAGGKFD

42. LIDEARTPLIISG

43. LLNRAPTLH

44. LPDKAIDLIDE

45. LPGKLADC

46. LSGGQQQR

47. MGHVDHGKT

48. NADFDGDQMAVH

49. NGAGKSTL

50. NLLGKRVD

51. NTDAEGRL

52. PSAVGYQPTLA

53. QRVAIARA

54. QRYKGLGEM

55. RDGLKPVHRR

56. SALDVSIQA

57. SGGLHGVG

58. SGSGKSSL

59. SGSGKSTL

60. SVFAGVGERTREGND

61. TGRTHQIRVH

62. TGVSGSGKS

63. TLSGGEAQRI

64. TNKYAEGYP

65. TPRSNPATY

66. VEGDSAGG

67. VRKRPGMYIG

50

In yet another embodiment of the present invention the number of invariant peptides varies according to the relatedness among the organisms and the number of organisms being compared.

In still another embodiment of the present invention the invariant sequences belong to following proteins as available in the database http://www.ncbi.nlm.nih.gov wherein the said list of proteins comprise:

I   DNA DIRECTED RNA POLYMERASE BETA CHAIN

II   EXCINUCLEASE ABC SUBUNIT A

III   EXCINUCLEASE ABC SUBUNIT B

IV   DNA GYRASE SUBUNIT B

V   ATP SYNTHASE BETA CHAIN

VI   S-ADENOSYLMETHIONINE SYNTHETASE

VII GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE

VIII ELONGATION FACTOR G (EF-G)

IX   ELONGATION FACTOR TU (EF-TU)

X   30S RIBOSOMAL PROTEIN S12

XI   50S RIBOSOMAL PROTEIN L12

XII 50S RIBOSOMAL PROTEIN L14

XIII VALYL tRNA SYNTHETASE (VALRS)

XIV CELL DIVISON PROTEIN FtSH HOMOLOG

XV DnaK PROTEIN (HSP70)

XVI GTP BINDING PROTEIN LepA

XVII TRANSPORTER

XVIII OLIGOPEPTIDE TRANSPORT ATP BINDING PROTEIN OPPF

In yet another embodiment of the present invention the said method of comparing the peptide libraries as given in step (iii) of claim 1 is carried out by following the steps given in figure 1.

In still another embodiment of the present invention the said method of locating the common peptides in the original protein sequences as given in step (iv) of claim 1 is carried out by following the steps given in figure 2.

In yet another embodiment of the present invention the said method of creating a common peptide of variable length after removing the overlappings as given in step (v) of claim 1 is carried out by following the steps given in figure 3.

One more embodiment of the present invention a microprocessor based system for performing the methods of the invention which comprises:

i) means of determining the amino acid sequence window for creation of peptide library and subsequent origin tagging,

ii) means of comparing the peptide library,

iii) locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location,

iv) joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

Another embodiment of the present invention, a computer based system for performing the methods of the invention further comprising a central processing unit, executing peptide library creating program (PEPLIB), peptide library matching program (PEPLIMP), peptide stitching program (PEPSTITCH), peptide extraction program (PEPXTRACT) wherein the said programs are all stored in a memory device accessed by the central processing unit connected to a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs with a user interface device.

In yet another embodiment of the present invention a method for assigning function to a protein of unknown function showing no/weak homology to other protein sequences in a publicly available database (SWISSPROT) by employing the following steps:

I.   generating computationally overlapping peptide library from the protein sequences of unknown function,

II.   sorting computationally the peptides of length 'N' (N is the length of the sliding window of amino acids) obtained as above, alphabetically, according to single letter amino acid code,

III.   matching computationally the current library with peptide library of all functionally known proteins to obtain common peptides,

IV. locating computationally these common peptides in the original proteins and subsequently labeling them with their origin and location,

V. joining computationally the overlapping common peptides to obtain a long chain of invariant peptide sequences,

VI. assigning function to the unknown protein based on the function of the protein with which maximum length of peptide sequence identity is found. The more is the number of matches with the proteins of similar function the likelihood of functional assignment will be higher.

The invention is explained with the help of the following examples and should not be construed to limit the scope of the present invention.

Example 1

Conversion of DNA sequence into alphanumeric sequence

The purpose of this module in our software is to translate computationally the whole query genome (DNA sequence) in all six reading frames using a specified codon table. Applicants used letter 'z' corresponding to the stop codons TTA, TAG and TGA, and letter 'b' for all triplets containing any non standard nucleotide(s) (K, N, W, R, and S etc.) while artificially translating the genome. Subsequently the translated genome sequence is converted computationally into an alphanumeric sequence ([0-9], 's', '*', and '-'.). Applicants search each overlapping heptapeptide in the peptide library, assign a corresponding number (occurrence value), and append it to the alphanumeric sequence. If a heptapeptide is not present in the library Applicants assign the number 0. If a heptapeptide begins with an amino acid corresponding to any of the start codon ATG,GTG and TTG Applicants append character 's' in the alphanumeric sequence. This will be helpful to detect the location of a probable start codon. In case a heptapeptide contains character 'z' Applicants append a character '*' corresponding to that heptapeptide. Thus consecutive seven '*' (*******) in the alphanumeric sequence is a signal for stop codon. Applicants append a '-' character for any heptapeptide containing character 'b'. This signals the presence of a non standard nucleotide character.

> PID 16127997   Homoserine Kinase (*E.coli*-K12)

53

*GTACCCTCTCATGGAAGTTAGGAGTCTGAC*ATGGTTAAAGTTTATGCCCCGGCT
TCCAGTGCCAATATGAGCGTCGGGTTTGATGTGCTCGGGGCGGCGGTGACACC
TGTTGATGGTGCATTGCTCGGAGATGTAGTCACGGTTGAGGCGGCAGAGACAT
TCAGTCTCAACAACCTCGGACGCTTTGCCGATAAGCTGCCGTCAGAACCACGG
GAAAATATCGTTTATCAGTGCTGGGAGCGTTTTTGCCAGGAACTGGGTAAGCA
AATTCCAGTGGCGATGACCCTGGAAAAGAATATGCCGATCGGTTCGGGCTTAG
GCTCCAGTGCCTGTTCGGTGGTCGCGGCGCTGATGGCGATGAATGAACACTGC
GGCAAGCCGCTTAATGACACTCGTTTGCTGGCTTTGATGGGCGAGCTGGAAGG
CCGTATCTCCGGCAGCATTCATTACGACAACGTGGCACCGTGTTTTCTCGGTGG
TATGCAGTTGATGATCGAAGAAAACGACATCATCAGCCAGCAAGTGCCAGGGT
TTGATGAGTGGCTGTGGGTGCTGGCGTATCCGGGGATTAAAGTCTCGACGGCA
GAAGCCAGGGCTATTTTACCGGCGCAGTATCGCCGCCAGGATTGCATTGCGCA
CGGGCGACATCTGGCAGGCTTCATTCACGCCTGCTATTCCCGTCAGCCTGAGCT
TGCCGCGAAGCTGATGAAAGATGTTATCGCTGAACCCTACCGTGAACGGTTAC
TGCCAGGCTTCCGGCAGGCGCGGCAGGCGGTCGCGGAAATCGGCGCGGTAGC
GAGCGGTATCTCCGGCTCCGGCCCGACCTTGTTCGCTCTGTGTGACAAGCCGG
AAACCGCCCAGCGCGTTGCCGACTGGTTGGGTAAGAACTACCTGCAAAATCAG
GAAGGTTTTGTTCATATTTGCCGGCTGGATACGGCGGGCGCACGAGTACTGGA
AAACTAA*ATGAAACTCTACAATCTGAAAGATCACAAC*

Computationally translated protein sequence

*VPSHGSzESD*MVKVYAPASSANMSVGFDVLGAAVTPVDGALLGDVVTVEAAETF
SLNNLGRFADKLPSEPRENIVYQCWERFCQELGKQIPVAMTLEKNMPIGSGLGSSA
CSVVAALMAMNEHCGKPLNDTRLLALMGELEGRISGSIHYDNVAPCFLGGMQLM
IEENDIISQQVPGFDEWLWVLAYPGIKVSTAEARAILPAQYRRQDCIAHGRHLAGFI
HACYSRQPELAAKLMKDVIAEPYRERLLPGFRQARQAVAEIGAVASGISGSGPTLF
ALCDKPETAQRVADWLGKNYLQNQEGFVHICRLDTAGARVLENz*MKLYNLKDH
N*

Computationally generated Alphanumeric sequence

********000*s01033000000s11124s0000s000050s100010000000000000001000000010000
00000000000000000s0s34444s869944000010s0000s0s000000000000s10ss223222444444

54

400s0000000s0ss1000000000s00000002s0111111333363211111100000001000002100000
000222000000s33330000000000000100010100001019411 1200s0000000000000000s000
00000211 1000100010*******s000000000

Example 2

Training of artificial neural network (ANN)

The purpose of this module in the software is to train the designed neural network (fig 2) with a specified no. of genes and non-genes. In this example the **training set** consists of 1610 *E.coli-k12* NCBI listed protein coding genes and 3000 *E.coli-k12* ORFs which have not been reported as genes (non-genes). The **validation set** has 1000 known genes and 1000 non-genes from *E.coli-k12*, distinct from those used in the training set. The **test set** contains another 1000 genes and 1000 non-genes from the same organism. For training of the ANN, genes and the non-genes are assigned a probability value of 1 and 0 respectively. To train the neural network, first Applicants convert all the *E.coli-k12* genes and non-genes into corresponding alphanumeric strings by the method described above (steps 2 and 3). Samples of two *E.coli-k12* genes and two non-genes in alphanumeric sequence format are shown in figure 3. Here it is important to note that the alphanumeric sequences corresponding to a gene is number rich compared to the alphanumeric sequences corresponding to non-genes. This supports our hypothesis. To quantify this number richness of an alphanumeric sequence, five parameters derived from the alphanumeric sequence have been selected. These five parameters are as follows:

*Total Score (*algebraic sum of all the integers of a given alphanumeric sequence), *Fraction of zeroes (*total no. of zero characters in the alphanumeric sequence divided by total no. of characters in the sequence), *Mean (* total score divided by total length of the sequence), *Variance (*variance of occurrence values about the mean occurrence value for the whole ORF), *Length of the maximum continuous non zero stretch* (represents the occupancy of uninterrupted non-zero numbers in a sequence)

Table 27(a): Training of ANN (genes)

S.No	Fraction of Zeros	Total Score	Average	Biggest Continuous stretch	Variance	Probability
1	0.663116	587	0.7816	19	2.10146	1

55

2	0.693950	214	0.7616	18	2.43068	1
3	0.597436	412	1.0590	13	3.16832	1
4	0.898876	12	0.1348	4	0.20654	1

Table 27(b): Training of ANN (Non-genes)

S.No	Fraction of Zeros	Total Score	Average	Biggest Continuous stretch	Variance	Probability
1	0.946429	3	0.0536	2	0.05070	0
2	1.000000	0	0.0000	0	0.00000	0
3	0.955556	2	0.0444	1	0.04247	0
4	0.956522	2	0.0435	1	0.04159	0

While calculating these parameters from the alphanumeric sequences characters's', '*' and '-' have been excluded. To determine the contribution of each parameter towards discriminating genes from non-genes, the neural network is trained using all the five parameters together. Parameters corresponding to alphanumeric sequences of genes and non-genes are calculated. The training, validation and test sets contain 6 columns, first 5 columns contains values of the 5 parameters and the last column contains the number '1' for genes and the number '0' for non-genes.

Example 3

The applicants have analyzed 10 prokaryotic genomes using the method of invention. Efficiency of the method has been defined as percentage of the NCBI listed protein coding regions predicted by said method. All the encapsulated protein coding regions have been eliminated automatically by a specifically developed program. The method is able to predict on an average 92.7% of the NCBI listed genes with a standard deviation of 2.8%. Both sensitivity and specificity values of the method are high except in M.Tuberculosis H37RV genome (as shown in figure No. 3).

56

<u>Example 4</u>

Prediction of start site of protein coding DNA sequences

Correct start site prediction rate of the method of invention varies from 49.5 % in M.Tuberculosis H37Rv (where specificity is also least) to 81.1 % in H.pylori 26695. The applicants method decides start location based on the presence of start codon plus conservation of the surrounding heptapeptides. This method can also be utilized to predict the start site of a query protein coding DNA sequences predicted by some other method. This can be done by simply converting the protein sequence into corresponding integer sequence and then deciding the valid start site 's' on the basis of surrounding heptapeptides. The applicants report three such cases from *E.coli* k-12 genome (two from the forward strand and one from the reverse strand), to exemplify the start site prediction (as shown below).

In prediction of start site there is a trade-off between number richness and length of the ORF. In Case 1(PID 16132273), the start location of the gene has been shifted from location 85540 to 85630 by NCBI. By visual inspection of the integer sequences corresponding to this gene it is evident that earlier there was a region after 's' which was full of zeroes; or in other terms not a number rich region (bold region in Case 1 of figure shown below). The start site has now been shifted so that it now lies before a number rich region as predicted by the said method of invention. Case 2 is an example of 5' upstream shifting of the start codon because there is a number rich region ('2011111' and one '3' and one '2') upstream of this start codon. So this has been shifted to location 4611050 from 4611194. Case 3 is another example of shifting of start site in the reverse strand where there is a number rich region ('16531311' and many other numbers in the string) upstream of the earlier NCBI start location.

Case1. PID 16132273

Location   Earlier NBI (85540 .....87354); New NCBI (85630......87354)

E

s0s0000000000000s000000000s00b.   4222s111000000000099992222421 0000s00s40004
466442223s0s0120000000177s9999985.. 3239888440s0011110001 13002s1116311112ss
22222s430100000000100s0100000639977. 0111001 0000000010000000000s2000010030
0000111101111 00000161171000000000s201s. 9000002ss10000000001099s76s621110
0s0s0000s00014444441111100000000000234331211000s033221s000000014s000s00000
0020000000000011 1000000000000000000000s000001s000000s48976531s1111111000.12234
59999999s92554010010s0s0002s2236667778s75221001s000s000ss00000066ss11111s32
111 00000s00000220433211000000000002100100 10000s00000s11000000354211s000000s
00s22*******

Case2. PID 16132266

Location   Earlier   NCBI   (4611194.....4611829);   New   NCBI
(4611050.....4611829)

N   E

s0002011111000000000000003000000000200000 10000030ss000000001110s0s000ss0000
0s102110000000100ss3s2000000000000000000000010002 1100011s110000000000s00000
000001s1010000001010000222222200000000000000000 10321002s3321111s1101111001
0000000s00s000s00101010100s00000*******

Case3. PID 16132224

Location   Earlier NCBI (2538824....2539273); New NCBI (2538824......2539699)
*******0000000000000ss000000001s200010422030000000 0s00000000000100000s0s98
889135120sss00012220000225 12s000022

E

3s123100000000ss0s000ss0022s30000ss00000s00000000000001000000s0s0000s1653131
1000000101010000s00200101s11

N

10000230ss0100000s0001000000s0000000s0000s0s00001100s001 10000000000000000s00
000s

E: Earlier start site at NCBI   ⟶   Forward reading frame

N: Newer start site at NCBI   ⟵   Reverse reading frame

Example 5

Prediction of protein coding DNA sequences

The method is utilized for prediction of protein coding DNA sequences for various genomes in a publicly available database (NCBI) by employing the following steps:

i) generating computationally overlapping peptide libraries from all the protein sequences of the

selected organisms available at http://www.ncbi.nlm.nih.gov,

ii) sorting computationally the peptides of length 'N' obtained as above, alphabetically, according to single letter amino acid code,

iii) cataloging every peptide and their unique occurrence different organisms,

iv) converting DNA sequence to alphanumeric sequence using peptide library obtained from steps 1 and 2,

v) retrieving all possible open reading frames (ORFs) from the alphanumeric sequence,

vi) training of the modified neural network for discriminating protein coding and non-coding DNA sequences,

vii) predicting DNA coding sequences in the open reading frames (obtained in step 4) using trained neural network,

viii) removing the encapsulated protein coding DNA sequences (genes within genes) .

Advantages:

1.   Main advantage of the present invention is to provide a new method for prediction of protein coding DNA sequences without using any external evidences like ribosome binding sites, promoter sequences, transcription start sites or codon usage biases.

2.   It provides a method for statistical analysis of protein coding DNA sequences that utilizes the biological information retained in the conserved peptides which withstood evolutionary pressure.

3.   It provides a simple method for start site prediction of a protein coding gene.

59

4. It provides a method to detect organism specific, strain specific protein coding DNA sequences.

5. It provides novel protein coding DNA sequences, which could be used as potential drug targets.

References:

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol Biol.*, **215**, 403-10

Bird,A.(1987) C$_p$G islands as gene markers in the vertebrate nucleus. *Trends Genet.*, **3**, 342-47

Chen,L., Ou,H., Zhang,R. and Zhang,C. (2003) ZCURVE_CoV: a new system to recognize protein coding genes in coronavirus, and its applications in analyzing SARS-CoV genomes. *Biochemical and Biophysical Research Communications*, **307**, 382-8.

Delcher,A.L. ,Harmon,D., Kasif,S., White,O. and Salzberg,S.L.(1999) Improved microbial gene identification with GLIMMER. *Nucleic Acid Research*, **27**, 4636-41

Kehoe,M.A., Kapur,V., Whatmore,A.M. and Musser,J.M. (1996)Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. *Trends Microbial.*, **4**, 436-43

Lukashin,A.V. and Borodovsky,M. (1998) GeneMark.hmm: New solution for gene finding. *Nucleic Acid Research*, **26**, 1107-15

Mathe,C., Sagot,M.F., Schiex,T. and Rouze,P. (2002) Current Methods of gene prediction their strength and weaknesses. *Nucleic Acid Research*, **30**, 4103-17

Medigue,C., Rose,M.,Viari,A. and Danchin,A. (1999) Detecting and Analyzing DNA Sequencing Errors:Toward a Higher Quality of the *Bacillus subtilis* Genome Sequence. *Genome Research*, 9, 1116-27

Pearson,W.R. (1995) Comparison of methods for searching protein sequence databases. *Protein Science*, **4**, 1145-60

Salzberg,S.L., Delcher,A.L., Kasif,S. and White,O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acid Research*, **26**, 544-8.

Shibuya,T. and Rigoutsos,I.(2002) Dictionary-driven prokaryotic gene finding. *Nucleic Acid Research*, **30**, 2710-25

Brahmachari, S.K.. and Dash, D. (2001) a computer based method for identifying peptides useful as drug targets. PCT international patent publication (WO 01/74130 A2, llth October 2001).

Chen,L, Ou,H., Zhang,R. and Zhang.C. (2003) Z-CURVE_CoV: a new system to recognize protein coding genes in coronavirus, and its applications in analyzing SARS-CoV genomes *Biochemical and Biophysical Research Communications,* 307,382-8.

Cumulative number of reported cases of severe acute respiratory syndrome (SARS) Geneva: World Health Organization, 2003. (Accessed April 9, 2003 at http://www, who.int/csr/sarscountry/2003_04_04/en/.)

Drosten,C., Giinther,S. and Preiser,W., (2003) Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N Engl J Med.,* (www.nejm.org on April 10,2003.)

Ksiazek,T.G., Dean Erdman,P.H. and Goldsmith,C.S. (2003) A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *NEnglJMed,* 348, 1947-58.

Lukashin., A.V. and Borodovsky, M.(1998) GeneMark.hmm New solution for gene finding *Nucleic Acid Research,* **26,** 1107-15.

Marra,M.A., Jones,S.J., Astell,C.R., Holt,R.A., Brooks-Wilson,A. (2003) The Genome sequence of the SARS-associated coronavirus. *Science,* **300,** 1399-404.

Mathe,C., Sagot,M.F., Schiex,T. and Rouze, P. (2002) Current Methods of gene prediction their strength and weaknesses. *Nucleic Acid Research,* 30, 4103-17.

Salzberg,S.L., Delcher,A.L., Kaif, S. and White.O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acid Research,* **26,** 544-8.

Tsang,K.W., Ho,P.L. and Ooi,G.C., (2003) A cluster of cases of severe acute respiratory syndrome in Hong Kong. *NEnglJMed,* **348,**1977-85.

Organization Applicant

       ,   Street : Rafi Marg
               City : New Delhi
               State : DELHI
               Country : INDIA
               PostalCode : 110 001
               PhoneNumber :
               FaxNumber :
               EmailAddress : ipmd@vsnl.net
       <110> OrganizationName : Council of Scientific and Industrial Research

Application Project

<120> Title : A computer based versatile method for identifying protein
coding DNA sequences useful as drug targets
<130> AppFileReference : 1729
<140> CurrentAppNumber :
<141> CurrentFilingDate : ____-__-__

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgccaattg aattaaaagt agaaggttta gtgggtaaac caaacgagaa aatttctgcg
60 gcagaatttc gtcaaaaatg tcgtgaatac gcggcggaac aggtcgaggg tcaaaagaaa
120 gactttatcc gtttaggtgt gttgggcgat tgggataatc catatctcac gatgaatttc
180 gataccgaag cgaatattat ccgcacttta ggtaaagtga ttgaaaatgg tcatttgtat
240 aaaggctcaa aaccagttca ctggtgtttg gattgcggtt cttctttagc agaagcagaa
300 gtggaatatg aagacaaagt ttctccgtca atttacgttc gtttccctgc ggaaagtgcg
360 gatgaaattg aagctaaatt ttctgcacaa ggtagaggac aaggtaaatt atcagccatc
420 atttggacta ccacaccttg gacgatgcca tctaaccgtg cgattgcggt gaatgcagac
480 ttagaataca acttagtcca acttggcgat gagcgtgtaa ttttagctgc tgaattagtt
540 gagtcagtgg caaaagcggt gggtattgag cacattgaaa ttctgggttc tgtaaaaggt
600 gatgatcttg aattaagccg tttccatcat ccgttctatg attttactgt gccagtgatt
660 ttaggcgatc acgtaaccac tgatggcggt acaggtttag tacataccgc acctgatcac
720 ggtttagacg actttatcgt gggtaaacaa tatgatttac caatggcggg tcttgtatcg
780 aatgatggta aatttatttc aacgaccgaa ttctttgcag gcaaaggcgt atttgaagca
840 aatccgcttg tgatagaaaa attacaagaa gtaggtaact tattaaaagt tgaaaaaatc
900 aaacacagct atccacactg ctggcgtcac aaaacgccaa ttatttttccg tgcaacaccg
960 caatggttta tcggcatgga aacgcaaggt ttacgccaac aagcattagg cgaaattaaa
1020 caagttcgct ggattccaga ttggggtcaa gcacgtattg agaaaatggt tgaaaaccgc
1080 ccagactggt gtatttcacg ccaacgtact tggggtgtgc cgatgacgtt gttcgtgcat
1140 aaagaaaccg aagaacttca tccgcgtacc ttggatctac ttgaagaagt ggcgaaacgt
1200 gtagaaagag cgggtattca agcgtggtgg gatttagacg aaaaagaatt attaggtgcg
1260 gacgcagaaa cctatcgcaa agtgccggat acccttgatg tatggtttga ctcaggatca
1320 acctattctt ctgttgttgc aaatcgccta gaatttaacg gtcaagatat tgatatgtat
1380 ttggaaggtt ccgaccaaca ccgtggttgg tttatgtctt ctttaatgct ttctaccgca
1440 acagacagca aagcaccata caagcaagta ttaactcatg gtttcactgt ggatggtcaa
1500 ggccgtaaga tgtcaaaatc tatcggtaac atcgtgacac cacaagaagt aatggataaa
1560 ttcggtggcg acattttacg tttatgggtt gcttctactg attatacggg tgaaatgacc
1620 gtttctgatg agatcttaaa acgtgcagcg gacagctatc gtcgtattcg taacaccgct
1680 cgtttcttat tagcaaactt gaatggtttt gatccaaaac gtgatttagt caaaccagaa
1740 aaaatgatta gcctagatcg ttgggcggta gcttgtgcat tagatgcaca aaatgaaatt
1800 aaagatgcat acgataacta tcaattccac actgtggtac aacgtttaat gcgtttctgt
1860 tcggtagaaa tggggtcatt ctatttagat attatcaaag accgtcaata tacaaccaaa
1920 gcagacagcc ttgcgcgtcg tagttgccaa actgcgttat ggcatattgc tgaagcatta
1980 gttcgttgga tggcaccgat cttatctttc actgccgatg aaatttggca cacttgcca
2040 caaactgaaa gtgctcgtgc ggaattcgta tttactgaag aattctatca aggcttattt
2100 ggtttaggcg aggatgaaaa attagatgat gcttactggc aacaactgat taaagttcgt
2160 tcagaagtga accgcgtatt agaaatttct cgtaacaata agaaatagg cggcggttta

2220 gaagcagaag tgaccgttta tgctaatgac gaatatcgtg cattattagc acaattaggc
2280 aatgaattac gttttgtatt aattacctca aaagtagatg taaaatcttt atcggaaaaa
2340 cctgctgatt tagcggacag cgaattagag ggtattgcgg taagcgtaac acgttctaac
2400 gcagaaaaat gccctcgttg ctggcattat tctgacgaaa ttggggtaag cccagaacat
2460 ccaacacttt gtgctcgttg tgtagaaaat gtagtgggca atggagaagt tcgatacttt
2520 gcataa
2526
<212> Type : DNA
<211> Length : 2526
      SequenceName : gi_GDC_HINF_1018846
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1018846

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttggaaaata aaatgacagt cgattacaaa aacactctta acctaccgga aaccagcttt
60 ccaatgcgcg gtgatttagc taagcgcgaa cctgataagt ag
102
<212> Type : DNA
<211> Length : 102
      SequenceName : gi_GDC_HINF_1021582
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1021582

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgaagataa ctcattgtaa attaaagaaa tctatacaaa ataagctact tgaatttttt
60 gtattagaag ttacagcccg agcagcggct gatttactcg atatctaa
108
<212> Type : DNA
<211> Length : 108
      SequenceName : gi_GDC_HINF_1082407
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1082407

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgtttctgg ttggaaacct tttgaggtgg gtttggcttg cgcttttat cattgcgcaa
60 atttgggctt atgtacaaac acctgattct tggttagcaa tgatttctgg tatttctggt
120 attttgtgtg tggtattggt aagtaaaggt aaaattagta attatttctt tggattgatt
180 tttgcctata cttattttta tgttgcttgg ggatcgaatt cttaggcga aatgaacacc
240 gtactttacg tatatttgcc ctctcaattt attggttact ttatgtggaa agccaatatg
300 caaaatagcg atggtggaga aagcgtgatt gcaaaagcgt taactgttaa aggatggatg
360 acattaattg ttgtgactac ggttggtact ttgctttttg ttcaagcatt acaagcggct
420 ggtggtagct caacaggttt agatggtcta actacaatta ttacggttgc ggcacagatt
480 ttaatgattt tgccgttatc gtga

63

504
<212> Type : DNA
<211> Length : 504
        SequenceName : gi_GDC_HINF_1144501
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1144501

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta ccttttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
<211> Length : 198
        SequenceName : gi_GDC_HINF_124181
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_124181

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgtttagtg gcgaacatga tgcttgcgat tgctatgtgg acctacaagc aggttctggc
60 ggcaccgaag ctcaagattg gacagaaatg ttgctccgta tgtatctccg ttgggctgaa
120 agcaaaggtt ttaaaacaga actgatggaa gtctctgacg gcgatgtagc tggattgaaa
180 tcagcaacca ttaaagtgag cggtgaatat gcttttggtt ggttacgaac agaaacgggg
240 attcatcgtt tagtgcgtaa aagtccattt gattccaata accgtcgtca cacatcattc
300 agcgcagcat ttgtctaccc tgaaattgat gatgatattg atattgaaat caatcctgct
360 gatttacgta ttgatgttta tcgtgcatca ggggcaggtg gtcagcacgt aaacaaaact
420 gaaagtgcgg tgcgaattac ccatatgcca agtggcattg tggtgcaatg tcaaaacgac
480 cgttcacagc acaagaacaa agatcaagca atgaaacaat taaaagcgaa attgtatgag
540 cttgaattac aaaagaaaaa tgcggataaa caagcaatgg aagataataa atctgacatt
600 ggttggggaa gccaaattcg ctcttatgta ttagacgatt cacgcattaa agatttacgt
660 actggcgtag aaaaccgtaa tacgcaagcc gtattagacg gggatttaga tcgatttatt
720 gaagcgagtt taaaagcggg cttgtag
747
<212> Type : DNA
<211> Length : 747
        SequenceName : gi_GDC_HINF_1279189
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1279189

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgcttggta acgaaaaaca agctgaagca caagctaaat atgcggaaga cacgctgaaa
60 caagcacgcg attttgctaa acaacatcat aaaacagcct atttagcgcg taatgcggat

6A

```
120 ggcttacaaa ctggtcaaaa aggttcgatt catacggaag caatggaatt ggttggcttg
180 gaaaacgtcg cagagggaga acaaaaaggc ttaactcaag tttcaatgga acagctttta
240 ttgtga
246
<212> Type : DNA
<211> Length : 246
       SequenceName : gi_GDC_HINF_1347200
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1347200

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgccacgta tttttgccgc ttgttttgtc ggggcggcgc ttgcttgtgg gggcgcaact
60 tatcaaggta tgtttaaaaa tccgcttgtt tcgccagata tttttgggtgt ttcagcgggg
120 gcaggttttg gggcaagttt ggcaattttt tataatttgc caatgattta tatccaattt
180 tttgctttta gcggtggcat tttagctgtg ttatgtgtat cgctcattgc ctcgcgtagt
240 cgtacacaag atcctatttt agtgctggtg ctttctggga ttgcaattgg ttctttactt
300 ggtgcaggca tttctttgtt aaaaattctt gcggatcctt tcactcaatt accttcaatc
360 acttttttggc tacttggtag cctgacggct attaatcaac aagatttaat tcaattgatc
420 ccgatgttgt tgctagggat tgttcccatt tttttattac ttactgatac gctggctcgc
480 acgattgcac cgattgaact gccactcggt attctgactt ctgcttgtgg ttattag
537
<212> Type : DNA
<211> Length : 537
       SequenceName : gi_GDC_HINF_1347942
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1347942

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgaagaact cattacggga gttaaaacnn gattatactg tggttatagt aactcataat
60 atgcaacaag ctacacgttg ctccgactat acggcattta tgtatttggg tgaattagtt
120 gaatttggtc aaacacaaca aatttttgat agacccaaga tacaacgtac agaagattat
180 attcgcggta aaatggggta g
201
<212> Type : DNA
<211> Length : 201
       SequenceName : gi_GDC_HINF_1476415
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1476415

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgattagtc tacaagaaac caaaatagct gtgcaaaatc taaatttcta ctatgaggat
60 tttcatgcat taaaaaacat taatttacgt atcgctaaga ataaagtgac cgcctttatt
120 ggtccttcag gttgcggtaa atctacttta ttgcggagtt ttaatcggat gtttgaacta
```

```
180 tatccaaatc aaaaagctac tggtgaaatt aatttagacg gtgaaaattt actcacaaca
240 aagatggata tttctctgat tcgtgctaag gttggtatgg ttttccaaaa accaacgcca
300 tttccaatgt cgatttatga taatattgca ttcggtgttc gtttgtttga aaaattatca
360 aaagaaaga tgaatgaacg agtagaatgg gcattgacta aggccgctct ttggaatgaa
420 gtgaaagata aattacataa aagcggagat agtttatctg cggacaaca gcaacgcttg
480 tgcattgctc gagggattgc tattaaacct agtgtgttgt gttagatga accttgttcg
540 gcattagatc ctatttcgac tatgaaaatt gaagaactca ttacgggagt taaaacnnga
600 ttatactgtg gttatagtaa ctcataa
627
<212> Type : DNA
<211> Length : 627
      SequenceName : gi_GDC_HINF_1476557
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1476557

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgagccagc ttaatattca atttccgaca aaattcaaac cgctctttga atctatttgg
60 cggtttatta ttttctacgg tgggcgaggt tcaggtaaaa gttttagtat cgctagagca
120 ttagtattgc gagcctatca atcgcctgtt cgagttttgt gttccgtgaa attcagaaat
180 cgatttctga ttctgtga
198
<212> Type : DNA
<211> Length : 198
      SequenceName : gi_GDC_HINF_1505851
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1505851

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtggttcccg agttcattat tgtttcttta atcttggtgg cacagtccat gaaattggcg
60 ttaaacaaat ggcttatcat atttggcaac gctatagctc ttcacataaa gtacgcttta
120 ttgcgattaa actttgaggg agttgttggt gagatttttag agaaagtcga taacggccaa
180 atgggcgttg tattaaaacg gatgatggtg cgagccgcaa gtaaagtcgc tcaacgtttc
240 aatattgaag caattgtgac aggggaggca ttagggcaag tttctagcca aactttaacc
300 aatttacgct tgattgatga agccgctgat gccttagtat tgcgtccgtt aattacccat
360 gataaagaac aaattatcgc gatggcgaaa gaaattggca ctgatgatat tgcaaaatct
420 atgccagaat tttgtggcgt gatttcaaaa aatcctacga ttaaagcggt tcgtgaaaag
480 attcttaaag aagaagggca ttttaatttt gagattcttg aaagtgcggt acaaaatgca
540 aaatatttag atattcgcca gattgcagaa gaaacagnaa aagcagtcgt ggaagtcgag
600 gcaatttctg tgttaggtga aaatgaagtg attttggata ttcgtagccc agaagaaacg
660 gatgaaaagc catttgaatc aggtacacat gacgtcattc aaatgccgtt ctacaaactt
720 tcttctcaat ttggtagcct tgatcaaagt aaaagttacg tgttgtattg tgaacgtggt
780 gtgatgagta aattacaagc cttatatttg aaagaaaatg gtttttcaaa tgtgcgtgta
840 tttgcaaaaa acattcatta a
861
<212> Type : DNA
<211> Length : 861
      SequenceName : gi_GDC_HINF_1524561
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_HINF_1524561

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttggccatcg ctattggtgg aggtaataga ggtaatgcaa gcggagtatt gcgccaaaat
60 tttgcagaag ataaagcaaa aaagaccgct tcgaagctcg tgggcgtaat ggctcactat
120 tttggcggta agtcgtttta tctgcccgca ggtgataaaa tcaaagaagc cttacgagat
180 gcacaaattt atcaagaatt caacggtaag aatgtacctg acctaataaa aaaataccga
240 ttgtcagaaa gcacaattta tgcgatctta cgcaatcaac gaacgcttca agaaagcga
300 catcagatgg attttaattt tagttag
327
<212> Type : DNA
<211> Length : 327
      SequenceName : gi_GDC_HINF_1568974
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1568974

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgtttaggt ggcactacct tggaggtttt acagtaatgc cagatacaaa taacacagaa
60 accaataata agatcgaact ctatctaaat ggcaaaattt tatccggttg gaaaagcctt
120 aacctgcaac gctcgctgga atcaatgagt ggtcgttttg atttaggcat tgctgtgcga
180 cctgaagatg atatatcagt gcttgccgca ggttcgccac tggtgctgaa aatgggcggg
240 caaaccgtga ttaccggtta cttggatgaa atcaaacaac gcgtaagcgg taacgacaaa
300 actatctctg tgagtggacg agataaaact tgcgacttgg tggattgtgc cattatccac
360 aacagctacc aattcaaaaa ccaaactgcc aaacaaattg ccgaagccat ctgtaaacct
420 tttggcatta gcgtagtatg gcaagtgcaa gcccctgaag ccaatgaacg aatccctgtc
480 tggcaagtag aaccaggcga aaccgccttt gataatttaa gcaaaatcgc ccgacacaaa
540 ggcgtgttag tcaccagcga cgtggacggc aatttgcttt tcaccgagcc gagcaacaag
600 caagtcggta atcttaccct tggcgaaaac ttgctcgaac tggaacaaac cgacagctgg
660 ttgcaacgct tttcgctcta tcgcgtgatt ggtgacgcag aacaaggcgg cgccaaaggt
720 gataccaaaa ccaaaaacaa agcggcaaaa ggcaaggaaa aagatgatgg cgtggtagaa
780 gatcccgata tttacccagg accagcagaa ggaggcaagt aa
822
<212> Type : DNA
<211> Length : 822
      SequenceName : gi_GDC_HINF_1586944
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1586944

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgaaggttt cttaccggct aaataattgt ctaagtttaa agttagcgct gatcccatta
60 ttaatactat tatttgttgt tatgggatcg gtgctttctt taatcgcaaa attagatttt
120 tattttttc aacaaatatt atttaattcc gaattgcatt ttgcattgct aatgtcattg
180 ggaacgtctc tttttctttt gatattagca ttatgtattg ctattccatc tgcatggcga
240 atgagtcaag tgcggttgcc ttttcaatca tttttgaca ctttgtttga tttaccaatg

67

```
300 gttttgccac cattagtcac aggactaagt ttgcttctac tttttagttc acaagggata
360 ttggctgaac tacttccttt tataagtaaa tggatttttt cccctgtagg gatcattatt
420 gctcagactt atattgcgag ttcgatttta ttgcgttgta gcgagccatt aaaactgcga
480 aaaaaaacca ttaaaactac gaaaataaaa ccttga
516
<212> Type : DNA
<211> Length : 516
      SequenceName : gi_GDC_HINF_1594339
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1594339

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgacaaaac gtaaaaatgt ttcctttact tatgaaaatt atactgttac gccattttgg
60 gatacgctca agttaagcta ttcacaacaa agaattacaa caagagcaag aacagaagat
120 tactgtgatg gtaatgaaaa atgtgactct tataagaatc cttagggct tcaattaaaa
180 gagggaaaag tcgttgatcg gaatggtgat cctgttgagt tgaagcttgt tgaggatgaa
240 caaggtcaga aacgacatca agttgttgat aaatataata atccttttag tgtagcctct
300 ggaactaata atgatgcttt cgtaggtaaa caattatctc cttctgagtt ttggttagat
360 tgctctattt taattgtga taagcctgtc agggtttata aatatcagta tagcaaccaa
420 gaaccagagt cgaaggaagt tgagttaaat agaaccatgg aaattaatgg aaagaaattt
480 gctacttatg agtctaataa ttatagagat agataccata tgattttacc aaaattctaaa
540 ggttacttgc ctttggatta taaagagcgt gatttaaata caaagacgaa acaaattaat
600 ttagatttaa caaaagcctt tactctcttt gagattgaaa atgaactttc ctatggtggt
660 gtttacgcga aaacgaccaa ggaaatggtg aataaagcag gatattatgg gcgtaatcct
720 acttggtggg cggagagaac gttagggaaa tcattgctta atggattgag aacgtgtaag
780 gaagattctt catataatgg gctactatgt cctcgtcatg aacctaaaac gtctttctta
840 attcctgtag aaacaacaac taagtctttta tattttgcag acaatatcaa gttgcacaat
900 atgttgagcg tagatttagg ttatcgttat gatgatatta aatatcagcc agagtatatt
960 cctggtgtaa cacctaagat tgcagatgat atggtcagag aattatttgt tccactccct
1020 ccagcgaatg gaaaagattg gcaaggaaac cctgtttata cacctgagca aattcgtaaa
1080 aatgcggagg aaaatattgc ttatattgca caagaaaaac gctttaagaa acattcttat
1140 tctcttgggg caacgttcga tcctctgaat tttttacgag tacaagtaaa atattcaaaa
1200 gggtttagaa ccccgacttc ggatgaactt tattttacct ttaagcatcc agattttacg
1260 attttaccca acccaaatat gaagccagaa gaggcaaaaa accaagaaat tgccttgact
1320 tttcatcatg attggggctt tttcagtaca aatgtatttc aaactaaata tcgccaattt
1380 attgatttag cttatctagg atcacgaaat ttatctaact ctgtggggtgg tcaggcgcaa
1440 gcaagggatt ttcaagtcta tcagaatgta aacgtagatc gtgcaaaagt gaaaggggtt
1500 gagattaact ctcgcttgaa tattggttat ttctttgaga agttagacgg ttttaatgta
1560 agttataagt ttacttatca aagaggacgt ttagatggta atcgaccaat gaatgcaatt
1620 caaccaaaaa cctctgttat tggattagga tatgatcata aagagcagag atttggagcg
1680 gacttatatg taacccatgt tagcgcgaag aaagctaaag atacttataa tatgttctat
1740 aaagaacagg gatataaaga tagtgctgtt cgttggagaa gtgatgacta tacgctagtt
1800 gattttgtta cttatataaa accagttaaa aatgtgactt tgcagtttgg tgtatataac
1860 ttgacagacc gtaagtattt aacttgggag tctgctcgtt caattaagcc atttggaaca
1920 agtaacttga ttaatcaggg aacaggtgcg ggtattaatc gtttctattc acctggtaga
1980 aactataaat tgagtgcaga aattacgttt taa
2013
<212> Type : DNA
<211> Length : 2013
      SequenceName : gi_GDC_HINF_1634710
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1634710
```

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgcgtgaac gtagttcgct ttctgctcta atggccaaaa cgattgaatg ggattttata
60 acagaaaacc ccctaaaata tcttgagaaa ccaaaagcgc cagcaccaag aactcgtcga
120 tataatgaac atgaaattga gcgtctgatt tttgtgtcag gttatgatgt cgaacatatt
180 gaaccgccaa aaaccttaca aaattgcacg ggggcggcat ttcttttttgc tatagagaca
240 gcaatgagag caggggaaat agcaagttta acttggaata atattaattt tgaaaagcgc
300 accacctttt tgccaattac taaaaatgga cattcacgca cggtgcctct ttcggtaaaa
360 gcaatagaga ttttacaaca tcttacttcg gtaaaaacag aaagtgatcc gcgagtattc
420 caaatggaag cacgccaact ggatcacaac ttccgcaagc tcaaaaagat ggaagggctt
480 gaaaatgcca atttacattt tcacgacacc cgccgtgaac gattggcaga aaaagtggat
540 gtaatggtat tagccaaaat atcgggccat agagatctca gtattctgca aaatacttat
600 tacgcacctg atatggcaga aggctataaa acaaaggcgg gttatgatct gaccccaacc
660 aaaggcttga gccaacggaa ttttttcttc tttaatgaaa acttcatcgt tttcacaaca
720 aatccaccga tagtcattaa gctgtaa
747
<212> Type : DNA
<211> Length : 747
     SequenceName : gi_GDC_HINF_1638626
     SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_1638626

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atggcgacaa ttatcaagaa tggcaagcgt tggcacgcac aagtgcgcaa gtttggcgtg
60 agcaaatcag ccattttttt gactcaagca gacgcaaaaa aatgggcaga aatgctcgaa
120 aaacagcttg aatcaggaaa gtataatgaa atccctgata ttacattgga tgaactcatt
180 gataagtatc taaaagaagt cactgtaacc aagcgcggga acgtgaaga gcgcataaga
240 ctactgcgtc tttctcgaac tccgcttgcc gcaatatctt tacaagaaat aggaaaagca
300 cactttcgtg agtggtaa
318
<212> Type : DNA
<211> Length : 318
     SequenceName : gi_GDC_HINF_1639409
     SequenceDescription :-

Custom Codon

Sequence Name : gi_GDC_HINF_1639409

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atggaagccg ttcaattaga caaaaatcaa gagcctaatt ataaaggtta tagcggtagc
60 ttgattcatc ctgcatttca acagcaaaca acaaaacgtg aaaaaccgag tacaccatta
120 cctagtttgg atttgctttt aaaatatccg ccaaatgaac aacgcattac accagatgaa
180 ataatggaaa cctcacagcg tattgaacaa caattacgca attttaatgt aaaagccagc
240 gtaaaagatg tgcttgttgg ccctgttgtt acgcgttatg aattagaatt acagccgggt
300 gtgaaagcat caaaagtcac gagcatcgat accgatttag caagagcatt gatgtttcgt
360 tctattcgtg tggcagaggt gattccaggt aaaccttata ttggtattga accccaaat
420 cttcatcgtc aaatggtgcc attacgtgat gtattagata gcaatgaatt ccgtgatagc
480 aaggcaactt tacctattgc tttaggtaaa gatattagtg gcaaaccagt cattgttgat

```
540 ttagcgaaaa tgccacattt attggtagca ggttctacgg gatcaggtaa gtctgttggt
600 gtgaatacga tgattctaag tttactttat cgtgttcaac cagaagatgt gaaatttatt
660 atgattgatc ctaaagtcgt cgaactttct gtttataatg atattccaca tttactgaca
720 ccagttgtaa cggatatgaa aaaagccgct aatgcgttgc gttggtgcgt agatgaaatg
780 gaacgtcgtt atcagttgct ttcagcttta cgcgtacgaa acattgaagg ctttaatgaa
840 aaaattgatg aatacgaagc aatgggaatg cctgtgccaa atccaatttg gcgactgggc
900 gatacgatgg atgcaatgcc accagcgttg aaaaaattga gttatattgt ggttattgtc
960 gatgagtttg ctgatttaat gatggtagcg ggtaagcaaa tcgaagaact gattgcacgg
1020 ttggcacaaa aagcacgagc tatcggtatc catttaattt tagccacaca acgcccctct
1080 gtggatgtga ttactggttt aattaaagca aatattccaa gtcgcattgc ctttacggtg
1140 gcaagtaaaa ttgactcacg tactattctt gatcaagggg gtgcagaagc ccttttaggg
1200 cgtggagata tgctttattc tggacaaggt tcatctgatt taatccgcgt acatggagcc
1260 tatatgagtg atgatgaagt catcaatatt gccgatgatt ggcgagcacg cggtaaacct
1320 gattatattg atggaatttt agaaagcgca gacgatgagg aaagttcaga aaaagggata
1380 tcaagcggtg gggaattaga tccactcttt gatgaagtaa tggactttgt tattaatact
1440 ggtacaactt cagtatcttc tattcaacgt aaattcagcg taggttttaa ccgagcagcg
1500 cgtattatgg atcaaatgga agaacaaggg attgtcagcc caatgcaaaa tggtaagcgt
1560 gaaattttat cgcatcgtcc agaatactaa
1590
<212> Type : DNA
<211> Length : 1590
       SequenceName : gi_GDC_HINF_1660491
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1660491


Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgtttatgc tttatttaga attttattt ttactattaa tgctctatat cggtagccgt
60 tacggcggta tcggattagg tgttgtttct ggtatcggtc ttgctatcga ggttttcgta
120 tttcgtatgc cagtggggaa gcaccgattg atgttatgct tatcattctt gcagtggtga
180
<212> Type : DNA
<211> Length : 180
       SequenceName : gi_GDC_HINF_170553
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_170553


Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgaataaaa tttttaaagt tatttggaat gttgtgactc aaacttgggt tgtggtgtct
60 gaactcactc gcgcccacac caaacgcacc tccgcaaccg tggcaaccgc cgtattggcg
120 accgtattgt ctgcaacggt tcaggcgatt aacgacgcag gaactttcgt gaaagtgcaa
180 agtacggaag atgatattga agatagtgct gcaaccaaag atgacaataa aaaccaagct
240 ctcaaagcag gcgacacctt aaccttaaaa gcgggtaaaa acttaaaagc taagttagac
300 caaggtggta aatcagtaac ctttgcttta gcgaaagacc ttgatgtgaa aaccgcgaaa
360 gtgagtgata ctttaacgat tggcgggaat acgcctgctg cgggtggtgc tacgccaaaa
420 gtaagtatta ctagcacggc tgatggcttg aagttagcaa aaggcactaa tggagatact
480 gcagttcatt tgaatggctt ggcttcaact ttgcctgatg tgactacaaa tacaggtgcc
540 tcaacttcag taaccttttc gcctagtgac attgaaaaaa caagagctgc aactattaaa
600 gatgttttaa atgcaggttg gaatattaaa ggagctaaag ttgcggggg taataccgag
660 aatgttgatt tagtggcggg ttatgacaat gttgagttta ttacaggaga taaaaacaca
```

```
720 cttgatgttg tattaacagc taaagaaaac ggtaaaacaa ccgaagtgaa gttcacaccg
780 aaaacttctg ttattaaaga taataatggt aagttgctta caggtaagca gttgaaggat
840 gcgaatactg gtacagcgac caatgcaact gaagatacag acgaggcaat ggcttag
897
<212> Type : DNA
<211> Length : 897
      SequenceName : gi_GDC_HINF_1807963
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1807963

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta ccttttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
<211> Length : 198
      SequenceName : gi_GDC_HINF_1817220
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_1817220

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atggctgctg caattcaaca acgtgccgaa cttcaacgcc gtatttggca aattgctaat
60 gatgtgcgag gctcggtcga tggctgggat ttcaaacaat atgtgcttgg cacacttttt
120 taccgtttta ttagcgaaaa ttttgccaat tacattgaag cgggcgatga aagcgtaaat
180 tatgcccaat tacctgatga aatcattaca cagatgccat aaaacgaaa ggctacttta
240 tttacccaag ccaattattt aagaatgttg cggctaatgc tggcagcaat cctaatttga
300
<212> Type : DNA
<211> Length : 300
      SequenceName : gi_GDC_HINF_231874
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_231874

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgaatactg atttaaaaca gattttact gatattgaaa actcagcgac gggctttccg
60 tctgaacaag atattaaagg gttatttgcc gattttgata ccaccagcaa tcgcttaggc
120 aataccgtaa aagataaaaa cgaccgctta acggctgttt tgaaaggcgt ggctgaactt
180 gattttggca aatttgaaga taaccacatt gatttatttg gcgatgcata cgaatatctt
240 atttctaact atgccgccaa tgcaggcaaa tctggtggcg aattttttac cccacaaagt
300 gtttccaaac tcattgctca aattgcaatg cacgggcaaa cctcggtcaa taaaatttat
360 gaccctgcag caggttctgg ctcacttttg cttcaagcca aaaaacaatt tgatgaacat
```

71

```
420 attattgaag aaggctttt cgggcaggaa attaaccata ccacatacaa ccttgcccgt
480 atgaatatgt ttttgcataa catcaactac gacaagtttg atattgcttt aggcaacacc
540 ttaatggaac cacaatttgg cgataataaa cctttcgatg ccattgtttc gaacccgcct
600 tactccgtga aatgggctgg ctccgacgat ccaacattga ttaatgatga acgatttgcc
660 ccccgcaggc gtgcttgcac caaaatccaa agcggacttt gcctttattt tacatgcgtt
720 aagttatctt tcagcaaaag gccgcgcggc gattgtttcc ttccctggta ttttttatcg
780 tggcggtgcc gagcaaaaaa ttcgtcaata tttggtggat aa
822
<212> Type : DNA
<211> Length : 822
      SequenceName : gi_GDC_HINF_232170
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_232170

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgatgaacg atttgccccc cgcaggcgtg cttgcaccaa aatccaaagc ggactttgcc
60 tttattttac atgcgttaag ttatctttca gcaaaaggcc gcgcggcgat tgtttccttc
120 cctggtattt tttatcgtgg cggtgccgag caaaaaattc gtcaatattt ggtggataat
180 aactatgtgg acgcggtgat tgcgcttgcg ccaaatctct tttttggcac cagtattgcg
240 gtgaatattt tggtgctttc caaacacaaa cccaatttat cgatgccagc ggtttattta
300 aatctgccac taataaccac attttag
327
<212> Type : DNA
<211> Length : 327
      SequenceName : gi_GDC_HINF_232813
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_232813

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgccgcatt tggcaaaatc catatccttt gaagaaatcg cccaaaatga ctacaacctt
60 gcagtaagtt cgtatgtgga acaaaaagac actcgtgaag tgattaatat tgatgaactc
120 aatgctcaaa ttcgtgaaac tgttaccaat attgaccact tgcgtgcgga aattgacaag
180 attgttgcag aaattgaagg gtaa
204
<212> Type : DNA
<211> Length : 204
      SequenceName : gi_GDC_HINF_233190
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_233190

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgacccaat acaaaactat cgctgaatcc aataattta tcgtttaga tcaatataat
60 aaatttgtgg aagaatctaa tgctggttat caaacggaaa ggagccttga gcgtgagttt
```

```
120 attcgtgatt tacaggctca aggctatgag tatttacaat ggcttaataa tcacgatgaa
180 ctgattaaaa acttacgggc gcaattacaa cgcttaaata acgtggtttt ctccgatgca
240 gaatggcaac gttttttaga ggaatatttg gataaaccga gcgataatct gattgagaaa
300 acccgcaaaa ttcacgatga ttatatttat gattttgtgt cgataacgg acgcattcag
360 aacatctatt tgcttgataa gaaaaatctt gccaataatt ctctgcaagt catcaatcaa
420 tttaagcaaa ctggcagcta tgataatcgt tatgatgtga caattttggt gaatggttta
480 ccccctttatt ga
492
<212> Type : DNA
<211> Length : 492
      SequenceName : gi_GDC_HINF_235441
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_235441

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atggtttacc cctttattga attaaaaaaa cgcggcgtgg cgattcgtga agcctttaac
60 caaattcacc gttacagcaa agaaagtttc aataaagaaa attctctctt taaatatatt
120 cagatttttg tcatttctaa tggcacggat actcgctatt ttgctaatac gactaaacgc
180 aataagaata gctacgactt cacaatgaat tgggcaacgg caaaaaatac tctgattaaa
240 gatttaaagg attttaccgc gactttcttg caaaagaata ctttgctcaa tgtgttggta
300 aattactgcg tgtttgatgt gagtgatacg ttgttaatta tgcgtccgta tcaaattgcc
360 gcaacagaac gtatttttatg gaaaattcaa atttcttact tagcaaaaaa ttggagtaat
420 cgtgaaagtg gtggctatat ttggcatacc acaggttcag gcaaaaccct caccagtttt
480 aaagcctctc gccttgcgac tgaacttgat tttattgata agtcttttt tgtggtcgat
540 cgtaaagact tagactacca aacgatgaaa gaatatcagc gttttttcgcc tgatagcgtg
600 aatgggtcgg aaagtaccgc tgggcttaaa cgcaatattg aaaaagatga taacaaaatt
660 atcgtaacca ccattcaaaa attgaataat ttaatgaaaa gtgaagaaaa cctgtctatt
720 tatcaaaaac aggtggtctt tattttcgat gaagcacatc gctctcaatt ggcgaagca
780 caaaaaaatc taaaacgtaa attcaaaaaa ttctatcaat ttggttttac tggcacgcct
840 attttccctg aaaacgcatt aggtgcggaa acgacagcaa gtgtgttcgg tgcggaattg
900 cattcttatg tgattaccga tgctattcgt gatgacaaag tactgaaatt caaagtcgat
960 tacaacgatg tccgcccaca atttaaagcc ttagaaacag aaaaagatcc tgaaaaattg
1020 accgcacttg agcagaaaca agccttttta caccctgagc gtattaaaga aatctcgcaa
1080 tatttgctta caatttttaa acagaaaacc caccgcttga atgccacagg taaaggcttt
1140 aatgcaatgt ttgcggtaag cagtgtagaa gcagcaaaac gttattacga aaccttacaa
1200 aatttacagg cagagcagga atatccgtta aaaattgcaa caatcttttc gtttgccgcc
1260 aatgaagaac aagatgcgat tggcgatatt ccagatgaga cctttgaacc cactgctcta
1320 aacagtacgg caaaagaatt tttaacaaaa gccattgatg attacaatca ctactttggc
1380 acgaattatg gcgtagatag tcaatcattc caaaattact atcgcgatct tgccaaacgt
1440 gtgaaaaatc aggaagtgga tttactgatt gtggtcggaa tgttcttaac aggctttgat
1500 gcccctacgt tgaatactct tttcgtggac aaaaacttgc gttatcacgg attaatgcag
1560 gcttttttctc gcaccaaccg catttatgac acaactaaaa cctttgggaa tatcgtcact
1620 ttccgcgatt tggaacaaaa taccattgat gcaatcacgc tatttggcga taaaaatacg
1680 aagaatgtcg tactagaaaa aagctacgac agctatttta cggtgatga caatcaacgt
1740 ggctatgcag aaatcgtgaa agaattaaaa gaaagctttc ctgatccaac agaaatcgaa
1800 acagagcaag ataaaaaaga gtttgtaaaa ttatttggcg aatatttgcg tgtcgaaaat
1860 attttgcaaa attacgatga gtttgcagca ttgcaagcct acaagcggt cgatttaaat
1920 gatccgattg caatggaaaa attcaagcaa gtacattatg tgaatgatga caaattgca
1980 gaaatgctga agtacccac tttaccagta gagcggagc aagattatcg ttcaacttat
2040 aatgatattc gcgattggtt acgccaaaga aaagagggga atgataagga taattcgccg
2100 ataaattggg atgatgtcgt gtttgaagtg gatttattaa aatctcaaga aatcaatttg
2160 gattatattc ttgcgttaat tttcgaacat cataagaaaa accaagacaa agaggtgtta
2220 attgatgaaa tccgccgcac agttcgttca agtttgggta accgtgcgaa agagagcttg
2280 attgtcgatt ttatcaacca aacaaattta gatgatattc ccgataaagc gactttaatt
2340 gactcattct tcctatttgc tcaagcagaa cagcgaaaag aagcagaatc cttaattcaa
```

73

```
2400 gaagaaaatt tgaatgtcga tgcagcaaaa cgctatatca gcacttcatt aaaacgggaa
2460 tatgccagtg aaaacggcac ggcacttaat gaagtattgc caaaaatgag tctacttaag
2520 ccacaatatc tcactaaaaa gcagaagatt ttccaaaaaa ttgctgcatt tgtagagaaa
2580 tttaaagggg ttggaggaaa gatttaa
2607
<212> Type : DNA
<211> Length : 2607
      SequenceName : gi_GDC_HINF_235913
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_HINF_235913

Sequence

```
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atggatataa taaagcctat atgcacaggt ttttttata acgataataa tgttttagga
60 gatttgatga aaaatttcaa atattttgct cagagttatg tggattgggt tattcgtctt
120 gggcgtcttc gtttttctct tttaggcgtg atgattctcg cggttttagc tctttgtact
180 cagattttat ttagtctatt tattgttcat cagatatctt gggtagatat ttttcgttcg
240 gtaacttttg cttactcac tgcgcctttt gttatttatt ttttcacttt attagtagaa
300 aaacttgaac attctcgtct tgatctttct agctcggtta atcgattgga aaatgaggtc
360 gccgagcgaa ttgctgctca gaaaaaatta tcccaagcat tggaaaagtt agaaaaaaat
420 agccgtgata aaagtacctt acttgccaca ataagccatg aatttcgcac gccattgaat
480 gggattgtcg ggcttagcca gatttttactt gatgatgaat tggatgatct ccagcgtaat
540 tatttaaaaa ctatcaacat aagtgcggtc agtttaggct atatttttag cgatattatt
600 gatttggaaa aaattgatgc cagccgaatt gaattaaatc gccagccaac agatttccct
660 gccttattaa acgatattta taatttttgct agtttcctcg ccaaagaaaa aaatcttatt
720 ttttctttag agcttgaacc taatttgcct aattggttga atcttgatcg tgttcgcttg
780 agccaaattt tgtggaactt aattagtaat gcggtgaagt ttacggatca gggaaatatt
840 attcttaaaa ttatgagaaa tcaggattgt taccatttta ttgtgaaaga tacaggaatg
900 gggatttcac ctgaagaaca aaaacatatt tttgaaatgt attatcaagt gaaagaaagc
960 cgccagcaaa gtgcgggtag cggtattggg ttggctattt ctaaaaatct tgctcagtta
1020 atgggaaggg gatttaacag ttga
1044
<212> Type : DNA
<211> Length : 1044
      SequenceName : gi_GDC_HINF_240336
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_HINF_240336

Sequence

```
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta ccttttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
<211> Length : 198.
      SequenceName : gi_GDC_HINF_243018
      SequenceDescription :
```

Custom Codon

74

```
------------
Sequence Name : gi_GDC_HINF_243018

Sequence
--------

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgaatattc atggtttagc aaaacttaat ggtaatgtca ctttaataga tcacagccaa
60 tttacattga gcaacaatgc cacccaaaca ggcaatatca aactttcaaa tcacgcaaat
120 gcaacggtaa ataatgccac gttaaacggc aatgtgcatt taacggattc tgctcaattt
180 tctttaaaaa acagccattt ttggcaccaa attcagggcg acaaagacac aacagtgacg
240 ttggaaaatg cgacttggac aatgcctagc gatactacat tgcagaattt aacgctaaat
300 aatagtactg ttacgttaaa ttcagcttat tcagctagct caaataatgc gccacgtcac
360 cgccgttcat tagagacgga aacaacgcca acatcggcag aacatcgttt caacacattg
420 acagtaaatg gtaaattgag cgggcaaggc acattccaat ttacttcatc tttatttggc
480 tataaaagcg ataaattaaa attatccaat gacgctgagg gcgattacac attatctgtt
540 cgcaacacag gcaaagaacc tgtgaccctt gagcaattaa ctttgattga aagcttagat
600 aataaaccgt tatcagataa gctcaaattt actttagaaa atgaccacgt tgatgcaggt
660 gcattacgtt ataaattagt gaagaataag ggcgaattcc gcttgcataa cccaataaaa
720 gagcaggaat tgctcaatga tttagtaaga gcagagcaag cagaacaaac attagaagcc
780 aaacaagttg aacagactgc tgaaaaacaa aaaagtaagg caaagcgcg gtcaagaaga
840 gcggtgttgt ctgataccc gtctgctcaa agcctgttaa acgcattaga agccaaacaa
900 gttgaacaga ctactgaaac acaaacaagt aagccaaaaa caaaaaaagg gcggtcaaaa
960 agagcattga gtgcagcgtt ttctgatacc ccgtttgatc taagccagtt aaaggtattc
1020 gaagtcaaac ttgaggttat taatgcccaa ccgcaagtga aaaagaacc tcaagatcaa
1080 gaggaacaag gcaaacaaaa agaattgatc agccgttact caaatagtgc gttatcggag
1140 ttgtctgcaa cagtaaatag tatgttttcc gttcaagatg aattggatcg tctttttgta
1200 gatcaagcac aatctgccct gtggacaaat atcgcacagg ataaaagacg ctatgattct
1260 gatgcgttcc gtgcttatca gcagaaaacg aacttgcgtc aaattggggt gcaaaaagcc
1320 ttagataatg gacgaattgg ggcggttttc tcgcatagcc gttcagataa tacctttgac
1380 gaacaggtta aaaatcacgc gacattaacg atgatgtcgg gttttgccca atatcaatgg
1440 ggcgatttac aatttggtgt aaatgtgggt gcgggaatta gtgcgagtaa aatggctgaa
1500 gaacaaagcc gaaaaattca tcgaaaagcg ataaattatg gagtgaatgc aagttatcag
1560 ttccgtttag ggcaattggg tattcagcct tatttgggtg ttaatcgata ttttattgaa
1620 cgtgaaaatt atcaatctga gaaagtgaaa gtgcaaacac cgagccttgc atttaatcgc
1680 tataatgctg gcattcgagt tgattataca tttaccccga caaataatat cagcgttaag
1740 ccttatttct ttgtcaatta tgttgatgtt tcaaacgcta cgtacaaac cactgtaaat
1800 agcacgatgt tgcaacaatc atttgggcgt tattggcaaa aagaagtggg attaaaggca
1860 gaaattttac atttccaact ttctgctttt atttctaaat ctcaaggttc gcaactcggc
1920 aaacagcaaa atgtgggcgt gaaattgggg tatcgttggt aa
1962
<212> Type : DNA
<211> Length : 1962
        SequenceName : gi_GDC_HINF_274892
        SequenceDescription :

Custom Codon
------------

Sequence Name : gi_GDC_HINF_274892

Sequence
--------

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgaaaaaaa ctgtatttcg tcttaatttt ttaaccgctt gtgtttcatt agggatagca
60 tcacaagcct gggcaggtca tacttatttt gggattgact accaatatta tcgtgatttt
120 gccgagaata aagggaagtt cacagttggg gctaaaaata ttgaggttta taacaaagaa
180 gggcaattag ttggcacatc aatgacaaaa gccccgatga ttgatttttc cgtggtgtcg
240 cgtaacggcg tggcggcatt agtaggcgat cagtatattg tgagcgtggc acataacggc
300 ggatataacg atgttgattt tggtgcagaa ggacgaaacc ctgatcagca ccgctttact
360 tatcaaattg taaaaagaaa taattatcaa gcttgggaga gaaagcatcc ttatgatgga
```

```
420 gattatcata tgcctcgttt acataaattt gtaactgaag ctgaacctgt gggtatgaca
480 acaaatatgg atggaaaagt atatgctgat agagagaact atcctgagcg tgtacgtata
540 ggctcaggac gtcagtattg gcgtacagat aaagatgaag aaacgaatgt acatagttca
600 tattatgtct caggtgcata tcgttatctt actgcaggaa atacccatac tcagagtgga
660 aatggtaatg gtacagtcaa tcttagtggt aatgtagtta gccctaatca ttatggtcca
720 ttaccaacgg gtggttctaa aggcgatagc ggttcgccaa tgtttattta tgatgcgaag
780 aagaaacaat ggcttataaa tgctgtatta caaactgggc atccttttt cggaagaggt
840 aatgggtttc agttaatacg tgaagaatgg ttttataatg aagttcttgc ggttgatacc
900 cctagtgttt ttcaacgcta tattccccca ataaatggac attattcctt tgtatcaaat
960 aatgatggta caggtaaatt aactttaact agacctagta aagatggctc taaagcaaaa
1020 tcagaagtag gaactgtgaa gttatttaat ccatcgttaa atcaaaccgc taaagaacat
1080 gttaaagcag cagcaggcta taatatttac caaccaagaa tggaatatgg aaaaaatatt
1140 taccttggcg accaaggaaa aggaacttta acaatcgaaa ataatataaa tcaaggtgct
1200 ggtggattat actttgaagg taattttgtt gtaaaaggca agcaaaataa tataacttgg
1260 caaggtgcag gcgtatctat tggacaagat gcaactgttg aatggaaagt tcacaatcct
1320 gaaaatgatc gtttatctaa aattggtata ggcactttat tagtcaatgg taagggaaag
1380 aatttaggaa gtttaagtgc gggtaacggc aaagtcattc tagatcaaca agcagatgaa
1440 gcgggtcaaa acaagctttt caaagaagtt ggcattgtaa gcggtcgagc aacagttcaa
1500 ttaaatagta cagatcaagt tgatcctaac aatatctatt cggatttcg tggtggtcgc
1560 ttagatctta acgggcattc attaaccttt aagcgtatcc aaaatacgga cgagggcgcg
1620 atgattgtga accataatac aactcaagtc gctaatatta ctattactgg gaacgaaagt
1680 attactgctc catctaataa aaagaatatt aataaacttg attacagcaa agaaattgcc
1740 tacaacggtt ggtttggcga aacagataaa aataaacaca atggacgatt aaaccttatt
1800 tataaaccaa ccacagaaga tcgtactttg ctactttcag gtggcacaaa cttaaaaggc
1860 gatattaccc aaacaaaagg taaactattt ttcagcggta gaccgacacc ccacgcctac
1920 aatcatttag acaaacgttg gtcagaaatg gaaggtatcc cacaaggcga aattgtgtgg
1980 gattacgatt ggatcaaccg tacatttaaa gctgaaaact tccaaattaa aggcggaagt
2040 gcggtggttt ctcgcaatgt ttcttcaatt gagggaaatt ggacagtcag caataatgca
2100 aatgccacat ttggtgttgt gccaaattaa
2130
<212> Type : DNA
<211> Length : 2130
     SequenceName : gi_GDC_HINF_276992
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_276992


Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgggggaaa acgcgatgaa tttaagtcgt cgagacttta tgaaagccaa tgcggctatg
60 gcagccgcaa cggcagcggg gctaaccatc ccagtcaaaa atgtggttgc ggctgaatcc
120 gaaattaaat gggacaaagc agtatgtcgt ttctgtggta ccggttgtgc agtattagtt
180 ggtactaaag atggacgtgt tgtggcatct caaggcgatc ctgatgcaga agtaaaccgt
240 ggtttaaact gtattaaagg ttatttcttg ccaaaaatta tgtacggtaa agaccgttta
300 acgcagccgc ttttacgtat gacaaacgga aaatttgata gaacggcga ttttgcgcca
360 gtttcttggg attttgccgt tcaaaacaat ggctga
396
<212> Type : DNA
<211> Length : 396
     SequenceName : gi_GDC_HINF_370413
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_370413


Sequence
```

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgataagaa cggcgatttt gcgccagttt cttgggattt tgccgttcaa aacaatggct
60 gaaaaattca aagaagcgtt caaaaagaac ggtcaaaatg cagtaggtat gtttagttct
120 ggtcagtcta ccatttggga aggctatgca aagaacaaac tttggaaagc aggttttcgt
180 tctaacaacg tagacccgaa tgcgcgtcac tgtatggcat ctgcagcggt tgcgtttatg
240 cgcaccttcg gtatggatga acctatgggt tgttataacg acattgaaca ggcagatgct
300 tttgttcttt ggggctcaaa tatggcggaa atgcacccaa ttttgtggtc gcgtattact
360 gatcgccgta tttctaatcc tgatgttcgt gtcactgtac tttctactta cgaacatcgt
420 agttttgaac ttgccgatca cggtttgata tttacaccgc aaactgattt ggcaattatg
480 aactacatca tcaattatct tattcaaaat aatgcgatta attgggattt tgttaataaa
540 cataccaaat ttaaacgcgg agaaacgaat attggctatg gtttgcgtcc agagcatcca
600 ttagaaaaag acacgaatcg taaaacagct gggaaaatgc acgattcttc ttttgaagaa
660 ttaaagcaac ttgtatcaga atatacagtg gaaaaagtat cgaaaatgtc tgggttagat
720 aaagtccagt tagaaacttt agcgaaactt tatgctgatc caacgaagaa agtggtttcc
780 tactggacaa tgggctttaa ccaacataca cgtggtgtgt gggtaaacca attaatctac
840 aatattcatt tacttactgg aaaaatttca atcccaggtt gtgggccatt ttcattaact
900 ggtcagcctt ctgcttgtgg tacggcgcgt gaagtaggtt cattccctca tcgtttacct
960 gccgacttag tggtaactaa tccgaaacac cgtgaaattg ctgaacgtat ttggaaatta
1020 ccaaaaggta cggtttctga aaaagtgggg ttacacacaa ttgcacaaga ccgtgcaatg
1080 aatgatggcg aaatgaatgt gttatggcaa atgtgtaaca ataatatgca agcagggcca
1140 aacattaatc aagagcgttt gccaggctgg cgtaaagaag caacttcgt gattgtttca
1200 gatccttatc caactgtatc cgcactttct gctgacttaa ttcttccaac ggcaatgtgg
1260 gtggaaaaag aaggggctta cggcaatgca gaacgtcgta cccaattctg gcgtcagcga
1320 gtcaaagcac cgggcgaagc aaaatcagac ttatggcaat taatggagtt cgcaaaatac
1380 ttcacgacgg atgaaatgtg gacagaagac ttacttgcac aaatgcctga atatagaggt
1440 aaaactttat atgaagtgct tttcaaaaat ggtcaagtag ataaattccc attgagcgaa
1500 cttgctgaag gccaattaaa cgatgaatca gaatattttg gttactacgt tcacaaaggt
1560 ttattgaag aatacgctga atttggtcgc ggtcacggtc acgatttagc tccatttgat
1620 atgtatcata agcgcgtgg tttacgttgg ccagttgtgg agggtaaaga aaccttatgg
1680 cgttaccgtg aaggctatga tccgtatgtt aaagaagggg aaggcgtggc tttctacggt
1740 tatcctgata agaaagcgat tattcttgcg gtgccttatg aaccaccagc tgagtctccc
1800 gataatgaat acgatttatg gttatcaaca ggtcgtgttc ttgaacattg gcatacgggg
1860 acaatgacac gccgtgtacc agaattgcat cgagcatttc caaataattt agtatggatg
1920 cacccattag atgcacaagc acgtggttta cgtcatggcg ataaaattaa gatttcatct
1980 cgtcgaggcg aaatgatttc ttatttagat actcgcggac gtaataaacc acctcgtggc
2040 ttagtattta ccactttctt tgatgcaggt cagcttgcaa atagcttaac actagatgcg
2100 acagacccaa tttcaaaaga aaccgacttc aaaaaatgtg cggtaaaagt ggaaaaggct
2160 gcgtaa
2166
<212> Type : DNA
<211> Length : 2166
       SequenceName : gi_GDC_HINF_370747
       SequenceDescription :

Custom Codon
Sequence Name : gi_GDC_HINF_370747

Sequence
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgttgttga aaggagtgat tatgcaggtc tcaagaagaa aattcttcaa gatctgtgca
60 ggaggtatgg cggaacgtc agctgcaatg ttgggctttg ctccagcaaa cgtattagct
120 gcgccacgcg aatataaatt attacgcgcg tttgaatccc gtaacacctg tacatattgc
180 gctgtaagtt gcggtatgtt gttatatagc acaggcaaac cttacaattc attaagcagc
240 catactggca caaatactcg ttcaaaactc tttcatattg agggtgatcc agatcatcca
300 gtcagtcgtg gtgcgctttg cccgaaaggt gctggctcac tcgattatgt caatagtgaa
360 agccgttctt tatatcctca atatcgtgcg ccaggttctg ataaatggga acgaatttct

```
420 tggaaagatg ccattaaacg tattgctcgt ttaatgaaag atgaccgaga tgccaacttt
480 gttgaaaaag attcaaatgg aaaaacggtt aatcgttggg caacgacagg aattatgact
540 gcatcagcaa tgagcaatga agctgcgtta ttaacacaaa agtggattag aatgctcggt
600 atggtgccag tatgtaacca agcgaatact tga
633
<212> Type : DNA
<211> Length : 633
        SequenceName : gi_GDC_HINF_5641
        SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_5641

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta cctttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
<211> Length : 198
        SequenceName : gi_GDC_HINF_628407
        SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_628407

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
atgacaaata actgggttga tattaaaaat gccaacttaa tcatcgttca aggcggtaac
60 cctgcagaag cccatcctgt tggcttccgt tgggcaattg aagcgaagaa aaacggtgcg
120 aaaatcatcg ttattgatcc gcgttttaac cgtacagcat ccgttgctga tcttcatgcg
180 ccaattcgtt ctggttctga tattacgttc ttaatgggcg tgatccgtta cctattggaa
240 acaaaccaaa ttcaacacga atatgttaaa cactatacca acgcatcatt cttaattgat
300 gaaggtttca aatttgaaga tggtttattt gtagggtata acgaagaaaa acgtaactac
360 gataaatcta aatggaacta ccaatttgat gaaaatggtc acgctaaacg tgatatgaca
420 ttacaacatc ctcgttgtgt cattaacatc ttaaaagagc acgtttctcg ttataccca
480 gaaatggttg aacgtattac aggcgtaaaa caaaaactct cttacaaat ctgtgaagaa
540 attggtaaaa cctctgtgcc aaataaaacg atgacgcatc tatatgcatt aggttttaca
600 gagcattcaa tcggtacaca aaatattcgc tcaatggcga taatccagtt acttttaggt
660 aatatgggga tgccaggtgg cggtattaac gcattacgtg acactccaa tgtgcaaggt
720 acgacagata tgggcttatt gccaatgtct ttaccaggtt atatgcgttt gccaaacgat
780 aaagatacct cttacgatca atacattaac gcaattacac aaaagatat cgttccaaac
840 caagtgaact attatcgtca tacttcaaaa ttctttgtta gcatgatgaa aactttctac
900 ggagataatg ccactaagga aaatggctgg ggattcgatt cttaccaaa agcagatcgc
960 ctatatgatc caattactca cgttaaattg atgaatgaag caaattaca cggttggatt
1020 ttacaaggtt ttaacgtatt aaattcacta ccaaataaaa ataaacgtt atctggtatg
1080 agtaaactga aatacttagt cgttatggat ccattacaaa ctgaatcatc agagttttgg
1140 agaaattttg gtgagtcaaa taatgtaaat cctgcagaaa ttcaaacaga agttttccgt
1200 ttaccaacta cttgtttcgc agaagaagaa ggatcaatcg ttaattctgg tcgctggact
1260 caatggcact ggaaaggttg cgatcaaccg ggagaagcct acctgatgt tgatattctt
1320 tctatgctac gcgaagaaat gcacgaactt tataaaaaag agggtggaca aggaattgaa
1380 tcttttgaag cgatgacttg gaattatgct caaccacact caccaagtgc ggttgaatta
1440 gccaaagaat taaatggtta tgcgcttgaa gatctttatg atccaaacgg taacttgatg
```

78

```
1500 tacaagaaag gtcaattact caatggattt gcacatttac gtgatgatgg tacaacaaca
1560 tcaggtaact ggttatatgt tggtcaatgg actgaaaaag gcaaccaaac tgctaatcgc
1620 gataattcag atccatcggg tttaggttgt actattggct ggggctttgc atggcctgca
1680 aaccgccgcg tactttatag ccgtgcatca ttagatatca atggtaatcc ttgggataaa
1740 aaccgccaat taatcaaatg gaacggtaaa aactggaact ggtttgatat tgctgactac
1800 ggtacgcaac caccaggttc tgatactggg ccgttcatta tgtccgcaga aggcgtagga
1860 cgtttatttg ccgttgataa aattgcaaat ggcccaatgc cagaacacta tgaaccagtt
1920 gaaagcccaa ttgatacaaa cccatttcat ccaaatgtag taaccgatcc aactttacgt
1980 atctataaag aagatcgtga atttattggt tcaaataaag agtatccatt tgtagcaaca
2040 acttatcgtt taaccgagca tttccacagc tggactgcac aatctgcatt aaatatcatc
2100 gcacaaccac aacaatttgt ggaaattggc gaaaaattag cggcagaaaa aggcatccaa
2160 aaaggcgata tggtaaaaat tacttctcgt cgtggctata ttaaagcggt cgccgtggtt
2220 acaaaacgtc ttaaagatct cgaaattgat gggcgtgtcg tacaccatat aggtcttcca
2280 attcactgga atatgaaggc attaaatggc aaaggtaacc gtggattctc tacgaatacc
2340 ttaacaccat cttggggtga ggcaatcacg caaacaccag aatacaaaac attcttggta
2400 aatattgaaa aagttgggga ggcataa
2427
<212> Type : DNA
<211> Length : 2427
     SequenceName : gi_GDC_HINF_6322
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_6322


Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttggttatgt tcaatgattt tttggcaaca ttcagccagc aattaacacc tcaaatgtgg
60 ggcgttgtcg caaccgcaac ttatgaaact gtttatatca gttttgcatc taccctactt
120 gctgtactag tcggcgtgcc tgttggcata tggacttttt taactggaaa aaatgagatt
180 ttacaaaata accgcactca ttttgtgtta aacacgatta ttaatattgg gcgttccatt
240 ccatttatta ttttgctcct aatcttatta cctgtaactc gtttcatcgt gggaactgta
300 ttaggtacaa cagcagcaat tattccattg agtatttgtg caatgccatt cgtggctcgc
360 ttaactgcta atgcactaat ggaaattcca aatggtttaa ccgaagcagc tcaagcaatg
420 ggggctacta aatggcaaat tgttcgtaaa ttctatttgt cagaagctct acctacgcta
480 attaatggcg ttactcttac gctagtcact ttagttggtt attctgcaat ggcaggaaca
540 caaggggcg gtggtttagg tagcctcgct atcaactacg ggcgtatatc gcaatatgcc
600 ttatgtaact ggggtggcaa ccattattat tgtgctattc gttatgatta g
651
<212> Type : DNA
<211> Length : 651
     SequenceName : gi_GDC_HINF_654365
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_654365


Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta cctttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
```

```
<211> Length : 198
      SequenceName : gi_GDC_HINF_661444
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_661444

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgcgtaaag atgcactacc cgcatttttc acagacgtaa atcaaatgta tgatgcctta
60 ttgaataaat caggggcaac aggtgtattt actgatttcc cagatacttg cgtggaattc
120 ttaaaaggaa taaaataa
138
<212> Type : DNA
<211> Length : 138
      SequenceName : gi_GDC_HINF_737160
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_737160

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtgatgagcc gacatcgagg tgccaaacac cgccgtcgat atgaactctt gggcggtatc
60 agcctgttat ccccggagta ccttttatcc gttgagcgat ggcccttcca ttcagaacca
120 ccggatcact atgacctact ttcgtacctg ctcgacttgt ctgtctcgca gttaagcttg
180 cttataccat tgcactaa
198
<212> Type : DNA
<211> Length : 198
      SequenceName : gi_GDC_HINF_775792
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HINF_775792

Sequence
--------
<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgcctaaac ctgaaccaat accacgaccg aggcgtttag cactatgctt tgcaccttca
60 gccggagata gagtatttaa acgcatctct tactcctcca ctttaaccat gtatgaaact
120 tggttaatca taccacgtac tgcaggcgta tcaattaact caacagtgtg gtgtatatgg
180 cgaagaccaa gaccacgcaa ggtagcttta tgcttcggta aacgagcaat tgagctacga
240 acttgtgtta ctttaatagt tttagccatt attcattacc ccaagatttc atcaacagtt
300 ttaccgcgtt ttgcagcaac catttctggt gatttcatat ttgctaatgc atcaatagtt
360 gcacgaacaa cgttaattgg gttggtagaa ccatacgctt agaaagaac gttacgtaca
420 cctgcaactt ccaataccgc acgcattgca ccaccagcga tgatacctgt accttcactt
480 gctggctgca taaatacacg tgaaccagta tga
513
<212> Type : DNA
<211> Length : 513
      SequenceName : gi_GDC_HINF_848166
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_HINF_848166

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
ttgtttatat atggggggaat aaatatgcaa attacacttt caaatacctt agcgaatgat
60 gcttggggaa aaaatgcgat tttgagcttt gactctaata aagctatgat tcatttaaaa
120 aataatggaa aaactgaccg cactttagtt caacaagctg ctcgtaaatt gcgtgggcaa
180 ggaatcaaag aggtggagtt ggtcggcgag aaatgggatt tggaattttg ctgggcgttt
240 tatcaaggtt tttataccgc aaaacaagat tacgcgattg agtttccaca tttagatgat
300 gaaccgcaag atgaattgtt agcacgtatt gaatgtggcg attttgtgcg tggaattatt
360 aatgaaccag cacaaagttt aacgcctgtg aaattagtag agcgagcggc tgaatttatc
420 ttaaaccaag cggacattta taatgaaaaa agtgcggtaa gtttttaagat tatttctggc
480 gaggaacttg agcaacaagg ttatcacgga atttggactg tgggtaaagg ctctgcgaac
540 ttgccagcca tgttgcaact tgatttcaat ccaacacagg attcgaatgc gcccgtgtta
600 gcttgtttag ttggtaaggg gattactttt gatagtggcg gctatagtat caaaccaagt
660 gatggtatga gtacaatgcg aactgatatg ggcggggctg cattattaac gggggcttta
720 ggtttcgcta tcgctcgtgg attaaatcaa cgcgttaagc tgtatttatg ttgcgcagaa
780 aatttggtaa gcaataatgc ctttaagcta ggcgatatta ttacttataa aaatggcgtg
840 agcgcagaag tactgaatac tgatgcggaa ggtcgtttgg tgttagctga tggattgatt
900 gaggctgata accaaaatcc aggtttttatt attgattgcg cgactttaac tggcgcagca
960 aaaagtggct gtaggaaacg actatcattc tgtattatct atggatga
1008
<212> Type : DNA
<211> Length : 1008
      SequenceName : gi_GDC_HINF_928073
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_928073

Sequence

<213> OrganismName : Haemophilus influenzae
<400> PreSequenceString :
gtggctgtag gaaacgacta tcattctgta ttatctatgg atgatgaact tgtgaaaaat
60 ctttccaat ccgcacaagc agaaaatgaa cctttctggc gtttaccatt tgaagatttt
120 catcgttcac aaaattaattc atcttttgcc gatattgcta atattggttc ggttccagtt
180 ggagctgggg caagcactgc aacggcattt ttatcgtatt ttgtaaaaaa ttataaacaa
240 aattggttgc atattgattg ctccgcgact tatcgtaaat ctggtagtga tttatggtct
300 gttggggcaa caggaattgg tgtgcaaact ttagctaatt taatgttatc aagatcattg
360 aagtaa
366
<212> Type : DNA
<211> Length : 366
      SequenceName : gi_GDC_HINF_929037
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HINF_929037

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :

```
gtgaaacaaa ttagtatctc ttgcagccat agaaaatatt ttgttagctt tagcgtggaa
60 tacgaacaag acattactcc cataaaaaac actaaaaatg gtgtggggct agatttgaat
120 atccttgata tagcttgttc ttgtgagata aacaaccatg acaaactaac ggactttaag
180 caataccaaa cagacatgaa agaattacta gggatagaaa tagatgaaga gctggatact
240 aaacgactta tccctactta ttccaaattg tattctttaa aaaaatactc taaaaaattt
300 aaaagattac aaagaaaaca aagccgtagg gtgttaaagt ctaaacaaaa caaaaccaaa
360 ttaggaggta attttacaa aacccaaaag aaattaaacc aagcctttga caagtctagt
420 catcaaaaaa cagacagata ccataaaatc acaagcgaac tttcaaagca atttgaattg
480 atagtagttg aagatttgca agtaaaaaac atgactaaaa gagctaaact caaaaatgtt
540 aaacaaaaga gtgggcttaa tcaatctatt ttaaacgctt cattctatca aatcatctct
600 tttttagact acaaacaaca gcataatggc aaattgttag tgaaagttcc cccacaatat
660 acgagtaaaa cttgccattg ttgtgggaat atcaaccaca agcttaaatt aaatcatagg
720 caatattggt gtttagaatg cgggtataga gaacacaggg acatcaacgc tgcgaacaac
780 attttaagca aagggttaag tcttttggg gtaggaaata tccatgcaga ctttaaagaa
840 caaagccttt cgtgttag
858
<212> Type : DNA
<211> Length : 858
     SequenceName : gi_GDC_HPYL_1068602
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_1068602


Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgaaagtca ataagggttt taaattccgc ttgtatccca ctaaagaaca acaagataag
60 ttgcaacact gctttttgt ctataatcaa gcttataata ttggcttgaa tgaactgcaa
120 gagcaatatg aaaccaacaa agattcacca cctaaagaaa gaaaatacaa aaaatcaagc
180 gaattagaca atgcgatcaa acaatgcttg agagctaggg acttgcccct tagcgctgtg
240 atagcccaac aagcacgcat gaatgttgaa agggctttaa aagatgcttt taaagttaaa
300 aacagaggct ttcctaaatt caaaaactct aaatccgcta aacaatcttt ttcgtggaac
360 aatcaaggct tctctatcaa agagagcgat gatgagtgct tcaagacatt cactctgatg
420 aaaatgcctt tactcatgcg catgcataga gacttccccc taattttaaa gtga
474
<212> Type : DNA
<211> Length : 474
     SequenceName : gi_GDC_HPYL_1069456
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_1069456


Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgatattca tcacccattt ttccacagag cctttacctt tacccatcct ggtttctaag
60 ggtttagcgg tcaaaggctt atcagggaat actctaatcc acaccttacc cgctcttta
120 atgtgccttg tcatggccac ccttgcggat tcaatttggc gtgaatcaat cctcccatgc
180 tctatggctt taatcgcaat atccccaaac gcaatggagt accccgatg ggctttccca
240 cgattgcgcc ctttcatttg ctttctgtat ·tttgttcttt ttggcattaa catgattatt
300 gcctccctct ctgcttcctt ctag
324
<212> Type : DNA
<211> Length : 324
     SequenceName : gi_GDC_HPYL_1376803
```

SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HPYL_1376803

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgagccgac atcgaggtgc caaacctccc cgtcgatgtg agctcttggg ggagatcagc
60 ctgttatccc cggggtacct tttatccttt gagcgatggc ccttccacac agaaccaccg
120 gatcactatg accgactttc gtctctgctt gacttgtatg tcttacagtc aggctggctt
180 gtgccattac actcaacttg cgatttccaa ccgcaatga
219
<212> Type : DNA
<211> Length : 219
      SequenceName : gi_GDC_HPYL_1474291
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HPYL_1474291

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgattaaac aaaccctcat cattcttgcc cctttttta tcgcaacgct gttgtatttt
60 ttaggcgcac cggatgggtt aagacctaac gcttggcttt attttgtat tttcatgggc
120 atgattatag ggctaatttt agagccggtg ccatcaggtt taatagcgct aagcgcgtta
180 gtgctgtgta tagcgttaaa aattggagcg agcgataaag tagcgagcgc taataaggct
240 atttcgtggg gtttgagcgg gtatgcgaat aaaacggtgt ggcttgtgtt tgtcgctttc
300 attttgggtt tagggtatga aaaaagcttg ttagggaaac ggatcgctct tttactgatt
360 aggttttttag ggcaaacccc tttaggttta ggctatgcga ttggtttgag cgaattgtgt
420 ctagccccctt ttatccctag caactccgct agaagtggag gcatactcta tcccatcgtt
480 tcatctatcc cgcctttaat gggatctact ccaaataata accctgacaa aatcggcgcg
540 tatttgatgt gggtcgcttt ggcttcaact tgcatcactt cgtccatgtt tttaaccgcg
600 ctcgctccta acccctagc aatggaaatc gctgccaaaa tgggcgtgaa tgaaatctca
660 tggttttcgt ggttttttagc gttcttgcct tgtggggtgg ttttgatctt gcttgtgcct
720 ttattggcgt ataaaacctg caaacccacc ttaaaaggct caaagaagt gagtttgtgg
780 gccaaaaaaa ggaattag
798
<212> Type : DNA
<211> Length : 798
      SequenceName : gi_GDC_HPYL_155367
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HPYL_155367

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgaacgccg catttaaaga aaggcgcttc attctcgtcc agttagatga aaaaattgat
60 cccaaggaag acaaaagcgc ttatgatttt tgtttgaaca ccttaaaatc accctcccca
120 agcatttttg acatcaccga agaaaggatt aaaagagcgg gggctaaaat caaagaagct
180 tgcgcgcatt tagatgtggg gtttagagcg tttgaaatca ttgatgatga aacgcatgct
240 aatgataaaa atctcagtca agcccatcaa aaggatttgt tcgcttattc taaccttgat

```
300 agaatggaaa cccaaacgat tttaattaag cttttaggct gcgagggttt ggagctcact
360 acccctataa cttgcttgat tgaaaacgcc ttgtatctgg ctttaaatac ggctttcatt
420 gtgggggata tagaaatgag cgaagtttta gaaaacttga aagataaagg ggtggaaaaa
480 atcagcatgt atatgcccgc tatcagtaac gataatttgt gtttggaatt gggcagtaat
540 ttgttggatt tgaaattaga gagtggcgat ttaaagatta gggggtag
588
<212> Type : DNA
<211> Length : 588
     SequenceName : gi_GDC_HPYL_1600102
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_1600102


Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgagccgac atcgaggtgc caaacctccc cgtcgatgtg agctcttggg ggagatcagc
60 ctgttatccc cggggtacct tttatccttt gagcgatggc ccttccacac agaaccaccg
120 gatcactatg accgactttc gtctctgctt gacttgtatg tcttacagtc aggctggctt
180 gtgccattac actcaacttg cgatttccaa ccgcaatga
219
<212> Type : DNA
<211> Length : 219
     SequenceName : gi_GDC_HPYL_447632
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_447632


Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
gtgcaacttc attgccacaa cttgccatgc gtttcaattg atattctact aggcggacca
60 ccatgccaga gctattctac ccttggcaaa agaaaaatgg atgaaaaagc gaatctgttt
120 aaagaatatt tgcggctttt agatttagta aaaccaaaaa tatttgtttt tgaaaatgtg
180 gtgggtttaa tgtctatgca aaaagggcaa ttattcaaac aaatttgtaa cgcttttaaa
240 gagagagatt atattttaga gcatgccatt ttgaacgccc tagattatgg tgtgcctcaa
300 atgagagaac gagtgatttt agtgggcgtg cttaaaagct ttaaacaaaa attttacttc
360 cctaaaccca taaaaacgca tttttctctg aaagacgctt taggggattt accacccatt
420 caaagcggtg aaaatggtga tgctttaggt tatcttaaaa atgcggataa tgttttttttg
480 gaatttgtgc gaaattctaa agaattaagc gaacatagca gtcctaaaaa caatgaaaaa
540 ctgataaaaa tcatgcaaac gctaaaagac ggacagagta aagatgattt gccagaaagt
600 ctgcgtccca aaagtggtta tattaatacc tatgccaaaa tgtggtggga aaaaccagcc
660 cccaccatta caagaaattt ttctaccca agcagttcta ggtgtatcca tccaagagac
720 tctagagcgt taagcattag agaggggca agattgcaaa gctttcctga taattataaa
780 ttctgtggga gtggtagcgc taaaagattg caaattggca atgccgtgcc gcctttattg
840 agtgtagcgc tcgcgcaggc ggtctttgac tttttaaagg ggtaa
885
<212> Type : DNA
<211> Length : 885
     SequenceName : gi_GDC_HPYL_506250
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_506250
```

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgtttgcag tgcatgctgc gatgattacg acattaaaga aagaagtttt ctttctttac
60 ctttatatca aatcactcaa aatcccgatt cctactacac tgaaatacat gatttcttta
120 ggcaaaatca gagaattaga tgttttagca aatcttgcta aactttgccc tacttgtcat
180 agggctttaa aaaaaggatc tagcgaagag gagtttcaaa aacgcttgat tagaaacatt
240 ctcaatcgca ataaagacaa tttagagttt gcgcaattgc gttttgaaac cgatgatttt
300 tcaacgctta ttgatcgtat ttgtgaaagc ttgaaatga
339
<212> Type : DNA
<211> Length : 339
     SequenceName : gi_GDC_HPYL_51094
     SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HPYL_51094

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgatggaat ttgatgttac catcatagat gagacaggca gggccacagc accagaaatc
60 ttgattcctg cacttcgcac taaaaaactg atcttaatag gcgatcacaa ccagctccca
120 cctagcattg ataggtacct cctagaacaa ttagagagcg atgatattca aaacttggat
180 gccattgatc gccaattatt ggaagagagt ttttttgaaa atctctataa gtatattcca
240 gagagtaata aggccatgct taatgagtaa
270
<212> Type : DNA
<211> Length : 270
     SequenceName : gi_GDC_HPYL_583607
     SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_HPYL_583607

Sequence

<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
atgcctgctt ctattggatc gctagttagt cagctttttt ataaagagaa acttaagaat
60 ggagtgatca aaaatacctc gcaattttac gatcctaaga atattatccg ttggattaat
120 gttgaagggg agcatcaact agaaaaaaca agtagctata acaaaaatca agttcaaaaa
180 atcatagagc ttttagagca aatcaatcgc gttcttaatc aaagaaaaat cagaaaaacc
240 ataggaatta tcacaccttta taatgcccaa aaaagatgct tgcgatcaga agtggaaaaa
300 tacggcttca agaattttga tgagctcaaa atagacactg tggatgcctt tcaaggcgag
360 aaggcagata ttattattta ttccaccgtg aaaacttatg gtaatctttc tttcttgata
420 gattctaaac gcttgaatgt agctattct agggcaaaag aaaatctcat ttttgtgggc
480 aaaaagtctt ctttgagaa tttgcgaagc gatgagaaga atatctttag cgctattttg
540 caagtctgta gatag
555
<212> Type : DNA
<211> Length : 555
     SequenceName : gi_GDC_HPYL_583883
     SequenceDescription :

Custom Codon

85

```
------------
Sequence Name : gi_GDC_HPYL_583883

Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgattattg aaacgcaaca agaccccaaa gaactacctg agtcttgcaa aataacgccc
60 caaaaaatct cttttaacca agtggttttt aaaaaaatta aaagaaaact caaccgcttc
120 attggaagca ttttagctcg gacagaagtg tataagaatc tcgtggcaaa atacgatgaa
180 ctcacaggaa aatacgaatc attàttggca aaagaggcaa acatcaaaga gaccttttgg
240 gaaaggcgtg ctgatagcga aaaagaagcc ttttttttag agcattttta cctcactagc
300 gtgtatgtgg cttctacagc aggatactat atcacgccta agggcgctaa aacctttata
360 gaagccacgg agcgtttaa aatcatagag ccggtggata tgttcataaa caaccccact
420 taccatgatg tggctaattt tacctatttg ccttgccctg tttcttaaa caagcatgct
480 ttcaatagca ccattcaaaa tgcaaaaaag cctgacattt cattaaaacc ccctagaaaa
540 tcctattttg ataatctttt ttatgatcaa ttaaacacta gaaagtgctt aaaagccttt
600 cacaaataca gcagacgata cgctcctta aaaaccccta aagaggttta a
651
<212> Type : DNA
<211> Length : 651
      SequenceName : gi_GDC_HPYL_665045
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_665045

Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgatggaaa ttttagtgtt gaatctgggc agttcgtcta ttaagtttaa gttgtttgac
60 atgaaagaaa ataagccctt agcgagcggt ttggctgaaa aaatcggcga agaaataggg
120 cagttgaaaa ttaaatcgca tttgcaccat aacgatcaag aattaaaaga aaagtttgtg
180 attaaagatc atgcgagcgg acttttaatg attcgtgaga atttaacgaa aatgggggatt
240 atcaaagatt ttaaccaaat tgacgctata gggcatcgtg tggttcaagg gggggataaa
300 ttccatgccc cagttctagt caatgaaaaa gtcatgcaag aaattggcaa tctttctatt
360 ttagccccct tacacaaccc ggcgaattta gccggtattg agtttgttca aaaagcgcac
420 ccccatatcc ctcaaatcgc tgttttgac accgcattcc atgccactat gcccagttac
480 gcttacatgt atgcgttacc ttatgaattg tatgaaaagt atcaaatccg gcactatggt
540 ttccatagga cttcacacca ttatgtggcc aaagaagcgg cgaagttttt gaataccgct
600 tatgaggaat ttaacgcgat cagtttgcat ttagggaacg gctcaagtgc agccgccatt
660 caaaagggta aaagcgtgga tacttctatg gggctaaccc ctttagaagg cttgattatg
720 ggcacaaggt gtggggatat tgaccccact gtggtggaat atactgcgca atgcgcgaac
780 aagagcttag aagaagtgat gaaaatgtta aaccatgaaa gcggattgaa aggcatttgt
840 ggggataatg agaaacatag aagccagaaa agaaaaaggt ga
882
<212> Type : DNA
<211> Length : 882
      SequenceName : gi_GDC_HPYL_953783
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_953783

Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
```

```
atgcctaaca gccaagtggc tgggcaagct agcgttttta ttttcccgga tttaaacgct
60 gggaacatcg cttataaagc ggtgcaacgg agcgctaaag ccgtggcgat agggcccatt
120 ttacaaggtt tgaataagcc cattaacgat ttgagtaggg gcgctttagt ggaagatatt
180 attaacaccg ttttgattag cgcccttcaa gcgcaagatt aa
222
<212> Type : DNA
<211> Length : 222
      SequenceName : gi_GDC_HPYL_954679
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_954679

Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
gtgagcctgg tttcaagcgt gtttttaatg tgtttagaca ctcaagtgct agtctttggg
60 gattgcgcga ttatccctaa ccctagccct aaagaattag ccgagatcgc taccacttcc
120 gcacaaaccg ccaagcaatt caatattgcg cctaaagtgg ccttgctttc ttatgcgaca
180 ggcgattccg ctcaaggcga aatgatagac aaaatcaacg aagctttaac aatcgctcaa
240 aagttggatc cccaattaga aattgatggc cccttacaat ttgacgcttc cattgataaa
300 agcgtagcca agaaaaaatg cctaacagcc aagtggctgg gcaagctagc gttttatttt
360 tcccggattt aa
372
<212> Type : DNA
<211> Length : 372
      SequenceName : gi_GDC_HPYL_954846
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_954846

Sequence
--------
<213> OrganismName : Helicobacter pylori
<400> PreSequenceString :
ttgaaagctg cacatcgttt gaatttaatg ggcgcggtag gattgatctt attaggcgat
60 aaagaagcca ttaattcgaa aaatttgaac ttgaatttag aaaatgtgga aatcattgat
120 cccaacactt ctcattatag agaagaattc gctaaaagct tgtatgaatt acgaaaatca
180 aagggcttga gtgagcaaga agctaagcaa ttagtgctgg ataagactta ttttgcgacc
240 atgctcgtgc attcaggcta tgtgcatgcg atggtttctg gggtgaatca cagctga
297
<212> Type : DNA
<211> Length : 297
      SequenceName : gi_GDC_HPYL_955261
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_HPYL_955261

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggtagcgg tccggattga agtcgtcggc catcgagtcc accacctggc cggccatctt
60 gagttccgcg ggtttgatct ccaccttctg gtccagcacc gggaagtcgg ggtcgcggat
120 ctcatcgggc cacagcaacg tgtgcaccat catcacctct cgcttgccga aatccttgac
```

```
180 gcgcaacgcc gccagcctgg tcttgttgcg cagcgtgaaa tgcacgatcg ccatccggtc
240 ggtctcggcg agtgtcttag ccagcagcac atacgatttc gacgacttcg aatcaggctc
300 caaaaagtag ctgcggtcga acatcatcgg gtccacgtcg gcggcgggga cgaactccaa
360 cacctcgatc tcccggctgc gttcttcagg caagctggcg atgtcgtcgt cggtgatcgc
420 caccatttgg ccgtcgccgg actcgtaggc ccgggcaaga tcgcggtagt cgaccacctc
480 gccacacgcc tcgcagacgc gcttgtaccg gatgcgtccg ttgtccttgg cgtgcacctg
540 gtggaacctg atgtcgtggt ctgcggtagc gctgtacacc ttgaccggca cgttcaccag
600 cccgaaggcg atcgaacccg tccaaatggc tcgcatgtaa gtgagtatgc cttgattgtc
660 cgcgagcgga acgtcacggc gaaattccac gcgatatttg accgtgacgt tacgctcgcg
720 acttgtgtga ccgacaggct acgttga
747
<212> Type : DNA
<211> Length : 747
        SequenceName : gi_GDC_MTUB_1045383
        SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1045383

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcgctcgg cgagggtgaa tccgccggcg cgcagtgcgg caagcacgcc atggtaccca
60 agcggatcgg tgaccaccgc cgcgctggga tggttttttgg cggcggcccg caccatcgcc
120 ggcccgccga tatcaatctg ctcgacgcag tcgtcgacac tggcgccgga ttcgacggtc
180 tggctgaacg gatacaagtt gactacaacg agttcgaaag cctcgatccc gagttgctcg
240 agggccgcgg cgtgctcgga cttgcgcagg tcagccagca gcccggcatg cactcgtggg
300 tgcagtgtct tgacccggcc atcgagcacc tcgggaaagc cggtcagctg ctccacgggg
360 gtcaccggaa tcccggtgtc ggcaatggtc ttggccgttg acccagtcga gatgatctcg
420 acgccggccg cgctcaggcc ctgtgccagg tctaccagcc cggtcttgtc gtacacgctg
480 atcagcgcac ggcggatcgg ccgtcttccg tcgtcggtgc tcatcctatg gttacctttc
540 gtcccatcgt cgctgttcgt ccgaccaccg tcacgccatg ggtggccagt gcggccaccg
600 ccgctaccaa cagccgtcgt tcggtga
627
<212> Type : DNA
<211> Length : 627
        SequenceName : gi_GDC_MTUB_1068100
        SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1068100

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcgcgctg acccgccgac gaccgcctgc aacacgcgat gcacgcccag cgtctgtgtc
60 ccgtcgatgt gcggtacatc gaccacctcg atgccgcccc gcagctgcgt cccggaaaaa
120 gtcaccttgc tgcagtcttt cccggggctg ggggccggca gcggctggga cgtctccacc
180 gcgatgacga cgaaccggtt gccgttgccc tcggcggaga cggcggccat gttgccctgc
240 aacccggtcg gcagctgggg cccggccgcc acttgcgcac agttcgccgg atcgaaactc
300 agcccgtcgg gcagtttgcg ggcggaaaag aacccgggat cgatggccct gggagtgaca
360 tcggtgacgg tgtattcagg tccaaagccc gacttcactt cggccacctt ggcgatgtcg
420 ccggtcgagg cggtggtgga gctggcccct gatgagcagc cgacaagcca gcacaccgat
480 ccgactgcca gtaccgcctt gcgcatcgtg gtcaatctac ccaacgcagc ccctgagctg
540 cgcaacgtcg acaccgtttt gactagcaga tcagcggcga actgcggtgc cagcggcgga
600 cgcaccgacc cggggtcggt gatcagccga cggcctcgat cacttgccgg ctacccggt
660 tga
```

88

663
<212> Type : DNA
<211> Length : 663
      SequenceName : gi_GDC_MTUB_1115707
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1115707

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgggtactg cgcaagagcg agtccgaagc cgatcaggcc cggttccgca ccacgctcta
60 cgtcacctgc gaggtagtcc gcatcgcggc actgctgatc cagccggtga tgccggagtc
120 ggccggcaaa attttggacc tgctcggcca ggccccaaac cagcggtcgt tcgccgccgt
180 aggtgttcgg ctgacccccg gcacagcgct gccgccgccc accggggtat ttccccgcta
240 ccagccgccg caaccacccg aaggcaagtg agcggaccgc agcgacggga aagccaccta
300 cgaagcgttg accgcggtct gcgcgtcgcg tgggatgtcg agcgtggcga cgggataaaa
360 cccggaatcg tcgcggccgt cgcgggacaa cagcatgggc ggatagttca ccacatggga
420 gccgttcggt ttgtgctgtt gccagtcgat cgcggcccgc agcgtgtagt ggcccgcggg
480 caagccggac agatcaacgc gaaccgtctc ggcgaccgac gccggtgtcg gctggtcgct
540 gctgcgatcg ccgcgctggt cggagaccag cgtcttcagg tccaccgctg ccggcagcgt
600 ccgaaccacc tgtccggtgg aatccaccag ccggtagccg ggcacccact tttcggtggc
660 ggcagcagcg ccgtagttgg tccaggtgac cgagatcgtc gcgaccttgc ccgctag
717
<212> Type : DNA
<211> Length : 717
      SequenceName : gi_GDC_MTUB_1124996
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1124996

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtcgatct ccggaatcga gcgctggtcg gctaccgaga acatccgcat ctcggtgatc
60 tcgtcgcccc agaactcgac ccgcaccgga tgttcggccg tcggggcaaa gatgtccaga
120 atcccgccgc gcacagcgaa ctcgccgcgc cggccgacca tatccacccg ggtatatgcc
180 agctcgacca gccgcgccac cacgccgtcg aaggggggatt cgtcgccaac ggtcagcgtg
240 aggggctcca tcatgcccag ctgcggcgtc atgggctgca gcagcgagcg caccgaggtc
300 accactaccc ccagcggtgg gcccagctgg gcatcgtcgg ggtgggccag ccggcgcagc
360 gccatcaggc gagtgccgac ggtgtcaaca ccgggtgaga gccgttcgtg cggcagtgtc
420 tcccaggacg gcaacaacgc caccgcatcc ccgaacacac cacgcagttc ggcggccagg
480 tcgtcggctt cccgcccggt ggcggtgacc accagcaatg gcccctgccg agccagcgca
540 ctggcgacca acagccgcgc gctggccggc gcgatgagcg tcaattcgtc gggtcgaccc
600 ccggcgcgct gcatgagctg ttggaatgtc ggcgcgctca gcgccaattc gacgagcccc
660 gcgatcgggg tatctgagca ggcaggcccc ggtgcggtca tgatgcggcc attctag
717
<212> Type : DNA
<211> Length : 717
      SequenceName : gi_GDC_MTUB_1138949
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1138949

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgctggcgt tctaccttcg gccaaggcca gggacgtggt gtacgagtga aggttcctcg
60 cgtgatcctt cgggtggcag tctaggtggt cagtgctggg gtgttggtgg tttgctgctt
120 ggcgggttct tcggtgctgg tcagtgctgc tcgggctcgg gtgaggacct cgaggcccag
180 gtagcgccgt ccttcgatcc attcgtcgtg ttgttcggcg aggacggctc cgacgaggcg
240 gatgatcgag cgcggtcgg ggaagatgcc cacgacgtcg gttcggcgtc gtacctctcg
300 gttgaggcgt cctgggggt tgttggacca gatttggcgc cagatctgct tggggaaggc
360 ggtgaacgcc agcaggtcgg tgcgggcggt gtcgaggtgc tcggccaccg cggggagttt
420 gtcggtcaga gcgtcgagta cccgatcata ttgggcaaca actga
465
<212> Type : DNA
<211> Length : 465
       SequenceName : gi_GDC_MTUB_1170285
       SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1170285

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgacgaccg ctggcataag cgggtcaaag ggccggacgg gaacaggcga accgtgcggt
60 ctgctgtctg cggcagggtt tcgcgctggc gcgtcaggtg ggttgacggc ggcggagagg
120 agcacagcaa gagcttccag cgcaaacctg acgcgcaggt acctgaccca tgccgaactg
180 ttgatgctcg ccagggccac gggccggttc gaaacgctca ccttggtgct cggctactgc
240 ggcttacggc ggtttacggt tcggtga
267
<212> Type : DNA
<211> Length : 267
       SequenceName : gi_GDC_MTUB_1176592
       SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1176592

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgggtcagt gcccacgacc tgtgcggcac tggccgcctg ccgtaattgt ttgtagccga
60 actaaattgc ggcgcgcctg cctgcgcgac taccgccgtc ccgccccctc cgacaagaag
120 cccaacaagt cgtaccgggt aatgacccca accggcttgc cttcctccac caccatcaac
180 gcatcccaat cacgcaacgc cttgccggcc gcactgacca attcaccggc gcctatcatc
240 cgcagcggcg ggctcatgtg tgccgacacg gcgtcggcca acttggcgcg gccctcgaac
300 acggccgaga gcagctcgcg ttccgagacg ctaccggcga cctcgccggc catcaccggc
360 ggctcggcgc cgaccaccgg catctgcgac accccgtact cgcgaagaat cccgatggcg
420 tcgcgcacgg tctccgacgg atgggtgtgc accagggcgg gcagcgcgcc ggacttgcgg
480 cgcaacacat caccgacggt ggattgctcg tcgacccgt caaggcggct gcgcaggaac
540 ccatag
546
<212> Type : DNA
<211> Length : 546
       SequenceName : gi_GDC_MTUB_1202653
       SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1202653

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttggcggcga tcccgagaag gtcacgctgt tcggtgaatc cgcgcgggaa tcgtcacgac
60 cctgctcgcc accccggcgg ccgcgggtct gttcgcggcg gcgatcgccc agagctcacc
120 ggcgacatcg gtctacgacc aggtgagggc tcggcgcgtc gcggtttgcg tcctcgacaa
180 gctgggaatc gacccgtccg atgtgcacag gttcatgaag tgccgaccgc ggcaatcctt
240 tccgcgtcca gcgaagtgtt caacgaagtg ccggttcgta accccggcac gctggcgttc
300 gtcccgatcg tcgacggcga tctgctgccc gactacccgg tcaagctggc gcaggagggc
360 cgctcacacc cggttccctt gatcatcggc accaacaagc acgagtcggc gctctttcgg
420 ttgatgcgct cgccgctgat gccgatcacc ccgcgcgatc acgtcgatgt tcacccagat
480 tgccgccgaa cagcccgatc tgcaagtgcc aaccgaggag cagatcggct ccgcgtactc
540 gcgatggcgg cgcaaagcac gctcattgag tatggctacc gacgtcggct tccggatgcc
600 gtcggtgtgg ctcgctga
618
<212> Type : DNA
<211> Length : 618
    SequenceName : gi_GDC_MTUB_1231843
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1231843

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgctggcct tgaggcccca gcgtcatttc acccagagcc ggagcgcccg gcggctacgc
60 tgtgtgctcg acgatgacgt atgggtgccc tgggcacggt caggggggttg caggacagca
120 acacggcatt tgtcggtgcg ctgcatagcg ggaacctgtt gggggccacc ggtgcggttc
180 tgcaggctcc gggcaacgcc gtcaacggtt tcttgttcgg ccagacgtcg atatcgcagt
240 cgattgacgt gtcaccggag tacggatacg agttggtcgc tgtcagcgac ccggttggcg
300 gaactgctgg ctccgctcga gccggtcacg gttacgttca cgccgacctt cggtgaaccg
360 gacatggtcc atctgagtgg cacgaagttc ggggccttg tcccggccct cttcgaaggg
420 gtgcgcgccg gcttctaa
438
<212> Type : DNA
<211> Length : 438
    SequenceName : gi_GDC_MTUB_1241031
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1241031

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgaccagct cagcaccgaa gcccgcggcg tcgcgcgcat cggactggcc aactacttcg
60 ccggcgcctt cctgctcccc taccgcgaat tccaccgtgc cgcagagcag ttacgctatg
120 acatcgacct gctgggccgc cggttcggag tgggcttcga aaccgtctgc caccggctct
180 ccacactgca gcgcccgcgg cagcgaggga taccgttcat cttcgtccgc accgacaagg
240 ccggaaacat ctcaaagcga cagtccgcga cggcgtttca cttcagccgg gtcggcggca

```
300 gctgcccgct gtgggtggtc cacgacgcgt tcgcccagcc agagaggatc gtccgccagg
360 tggcgcaaat gcccgacggc aggtcgtact tctgggtggc caagaccacc gctgccgacg
420 ggctcgggta tctgggcccg cacaagaact tcgcggtcgg gctgggctgc gacctcgcgc
480 acgcccataa actcgtctac tccaccggtg tcgtcctgga cgacccgagc acggaggtcc
540 cgatcggggc gggtcgcaag atctgcaacc gaacgtcgtg cgcccaacgt gcgttcccct
600 atctcggtgg tcgcgtcgcg gtcgacgaga acgcgggcag cagcttgcct tattcgtcga
660 ccgagcaatc ggtttgaccg cccgacgcca cagcagacaa cgaaacccct tatattactg
720 tggtttcagc aggctctggg caagcattgt tgtcggtgcc tgcacatagc attcagtcat
780 gtgttccact cgggaggaga tcacggaggc cttcgcgtca ttggctaccg cgctgtcccg
840 cgtgctgggg ctgacctttg a
861
<212> Type : DNA
<211> Length : 861
        SequenceName : gi_GDC_MTUB_1252888
        SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_1252888

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcagcttg gcaatcaaaa cactatgaga ttcgcagggc ggcctcagcg ttttcgccaa
60 agcgcttacc ccctgttcaa ccccaacagc gcgatcgcgc ttggccaccc attcggcggc
120 tcgggggcac ggttgatgac tacagtgcta caccacatgc cggacaaggg aattcgctac
180 ggcttacaga cgatgtgcga gggccgcggc caagccaatg ccaccattgt ggagttgctg
240 tga
243
<212> Type : DNA
<211> Length : 243
        SequenceName : gi_GDC_MTUB_1264312
        SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_1264312

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgacggtat accgtcgagg tatggctgtg ttaacggatg agcaggtcga cgccgcactg
60 cacgacctca acggctggca gcgcgccggt ggtgtcctgc gtaggtcaat caagtttccg
120 acgtttatgg ccggtatcga cgccgtacgc cgggtggccg agcgagccga ggaggtaaat
180 catcatccgg acatcgatat ccgttggcga acagtaactt tcgcgctggt tacgcatgcg
240 gtaggtggta tcacggaaaa cgacattgcg atggcgcacg atatcgacgc aatgtttggg
300 gcctaa
306
<212> Type : DNA
<211> Length : 306
        SequenceName : gi_GDC_MTUB_1286282
        SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_1286282

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgggtgcag tacggcttca acctcaccgc atgggcggtg ggatggctgc cctacatcgg
60 catactggca ccgcagatca acttcttcta ttacctcggc gagcccatcg tgcaggcagt
120 cctgttcaat gcgatcgact tcgtggacgg gacagtcact ttcagccagg cactaaccaa
180 tatcgaaacg gccaccgcgg catcgatcaa ccaattcatc aacaccgaga tcaactggat
240 acgcggcttc ctgccgccgt tgccgccaat cagcccgccg ggattcccgt ctttgcccta
300 acttcggact ag
312
<212> Type : DNA
<211> Length : 312
      SequenceName : gi_GDC_MTUB_1301742
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1301742

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgccttcgc cggtgagcag cggaccgacc agccatggca caaacaaggg gtgcgggttg
60 atcaggtctg agtcgatgaa caccacgatg tcgccgctgg tggccgccag tgaacgccac
120 aatgcctcac ctttgccggg ccgtaccggc acctcgggca acgcctgttc acggctgaca
180 acccgggcgc cggaggcgat ggcccggatc tcggtgtcgt cggtggaacc ggagtccagc
240 acgatcaatt catcgaccag gccatcgacc agcggagaga tgctgtcgat caccgattcg
300 atggtcgctt cctcgttgag ggccggcagc accaccgaaa tcgtccgtcc ggcctttgcc
360 gcttccaact ccccgatcgt ccagccggga cggtgccaag tagtgtccaa gggcagcgcg
420 ccaggggccc tgccaccggc gagatcgccg gcgaccagct ccgatgctgt catgcgagtc
480 ctctcaccgt gcgcgtcggc ggccggaccc cctgaatcga tgccaccatt ccagcaccc
540 gccgggtggc ggcgacctca tgcacccgaa acatgcgcgc cccggcggcc gcagccaacg
600 cggtggctgc cagcgttccc tcaagccgtt cggtcaaatc cacgcccaga gtctccccga
660 caacgtcctt gttgctcaaa gccatcagca cgggccaccc ggtcataa
708
<212> Type : DNA
<211> Length : 708
      SequenceName : gi_GDC_MTUB_1351907
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1351907

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgctttcag cggttatcct gaccgaacgt ggctatccag cggtgcccct ggcgggacaa
60 ctggtgcacc agaggttcgt ccgtcccggt cctctcgtac tagggacagg tttcctcaag
120 tttctgacgc gcgcggcgga tagagaccga actgtctcac gacgttctaa acccagctcg
180 cgtgccgctt taatgggcga acagcccaac ccttgggacc tgctccagcc ccaggatgcg
240 acgagccgac atcgaggtgc caaaccatcc cgtcgatatg gactcttggg gaagatcagc
300 ctgttatccc cggggtacct tttatccgtt gagcgacacc ccttccactc gggggtgccg
360 gatcactaa
369
<212> Type : DNA
<211> Length : 369
      SequenceName : gi_GDC_MTUB_1476279
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1476279

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttggtgggac gcagccgcgt actcgtcctg ttcggagcgg gtgaacatgt cgacgtcgtt
60 gcgttgctcg gtgagcgcgc ccatcggctg atcggtgaac acgtcgtgca gaccgtcgta
120 ggccatgtgg tccaaaaccg taacgtcgcc gtacttgtaa cccgaccggc tattcatcaa
180 caggtggggc gccttcgtca tcgactcctg accgccggcc accaccacgt cgaactctct
240 ggcccgaatg agttgatcag ccagcgcgat tgcgtcgatg ccggacaggc acatcttgtt
300 gatcgtcagc gcagggacat cccaaccgat gccggccgcc actgccgcct gccgtgcggg
360 catttgcccg gcacccgcgg tcaacacctg gcccatgatc acgtactcga ccaaggacgc
420 cggcacgttg gccttctcca gggcgccctt aatggcgatg gcacccagct cgctggcgct
480 gaaatccttc agggagccca tcaacttgcc gatgggtgta cgcgcgccag caacaatcac
540 cgatgtcgtt atgactacct cctcagcgca cccgaaagcc gatctgaccg acccggagaa
600 gcagattctt tcccttcagg ttaccgttgt gtgatgacga ccgatcaagt ccacgcccgt
660 cacatgctgg ctacctcgtt ggtaactgga ctcgatcacg tcggtattgc ggtcgccgac
720 ctggacgttg ccatcgagtg gtatcacgac caccttggca tgatcctggt ccacgaggaa
780 atcaacgacg atcagggcat ccgcgaggca ctgctggcgg tgccgggctc cgcggcgcaa
840 atccagttga tggccccgct cgacgaatcc tcggtgatag cgaagttcct ggacaagcgc
900 gggccaggca tccaacagct ggcgtgccgg gtcagcgatc ttgacgccat gtgtcggcgg
960 ctgcgctccc agggcgtccg gctggtctac gagacggcca ggcgtggcac cgcgaactca
1020 cggatcaact tcatccaccc gaaagacgcc ggcggggttc tgatcgagtt ggtggagccg
1080 gcccctaa
1089
<212> Type : DNA
<211> Length : 1089
        SequenceName : gi_GDC_MTUB_1485311
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1485311

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcgcgcgg caacaaagtc gccatcctcg agctgctggc gcgcctgtgc caccgctgga
60 tcgacttcgg tggactcctc ggaactcgct gcgcccttga gctttccggc tgtcgcagac
120 aacagggaat ccacccagcg actcagttgg tccgcgggct ggaggccctg gaagctcgag
180 atcggctgtc ccgcagccaa ggccaccacg tcggaaccg cttggacgcc gaatatctgt
240 gccaccctgg gtgcgacgtc aacgttaacc gacgccagcg accacttgcc cttagcggca
300 gcggccaagc cggacagcgt gtcaagcaag tcgacgcata cctcgctgcg gggtgaccac
360 agcaacacca ccaccggcac ttcgtcggac cggacgatca cctcgtcctc gaagttcgcc
420 tcggtgatct cggtcacacc ggacggcgtc gacagtgccc ggtcggcatc cgtgctcgcc
480 gcagcgtttt gctgggcacg ttgtttgatg ccggagaggt caacagcacc ggccatggcc
540 ggcccgagcg ggggtcgcgg acgcgtcacg ccgtcaagtc tgtcatgccg ctgcggtcat
600 cgatccaccc ggtggcgccg accctgcggc aggagccgac ataccgcgat cggttggtat
660 gaccaagatc acactggccg ccaccgaccc ctcaaccgct atccggcccg caatatcagt
720 gcgtcgccct gcccgccagc cccgcacaat gcggcaaccc gacgcccga tccccggcgt
780 gccaactgca cgccgcatg tagcgtgatt cgcgtccctg acatgccgag gggatgcccg
840 acggcaatcg caccaccgtt gacgttgacg atctgggggt tcagcccgag ttcgcgtatc
900 gaggccaatg ccaccgcagc gaacgcctcg ttgatctcca ccacgtcgag ctggtccacc
960 gagatgccct cgcgatccag cgccttgttg atcgcgttgg ccggctgcga ttgcagtgtg
1020 gaatccggcc cggccaccac accgtgggcg ccgatctcgg ctagccaggt cagccccagt
1080 tcctgggcct tttcctggtt catgaccacc accgcggccg caccgtcgga gatctgtgac
1140 gccgacccgg cggtgatggt gccgtcgcca cggaacgccg gcttcagacc ggccagcgcg

```
1200  gcggcggtgg  tgttggcgcg  gatcccctcg  tcctcggtga  actgcagtgg  atcgcccgtg
1260  cgctgcggga  tgttcaccgg  gatcacctcg  tcggcgaata  cgccgtcctt  ccatgccgcg
1320  gccgcctttt  ggtgggacgc  agccgcgtac  tcgtcctgtt  cggagcgggt  gaacatgtcg
1380  acgtcgttgc  gttgctcggt  gagcgcgccc  atcggctga
1419
<212> Type : DNA
<211> Length : 1419
      SequenceName : gi_GDC_MTUB_1486309
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1486309

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcggtcac  ggcgtctagc  acccaccgg  ccacggtcgc  ggcggacagc  cagcccagcc
60   acagccacgc  gcgctgcggc  gcctccccga  acaacgccgc  catcagcggc  accagcaaca
120  cggtgcccac  cgctcgcgcg  acaacggaac  aaaacgcgag  cagcgcaaag  ccgattagcc
180  tggcgcggtg  gtcgttcgga  acaagggcta  tccaggtgcg  gatcatcggg  tgccgtcctg
240  cgctgcggcg  accgccaccc  ggctgccctg  gccggtgtcc  cacagccggc  agtagcgtcc
300  gcccgcggca  agcaactcct  cgtgggtgcc  gcgttcgacg  atccgaccat  gatcgagcac
360  gacgatctgg  tcggcccggg  tgatggtatg  cagtcgatgg  gcgattacca  gcacggtgcg
420  gtcccgggtc  agccggttaa  gcgcctgttg  cacaaggtat  tccgattccg  gatcggcaaa
480  cgcggtggcc  tcgtcgagga  tgaggaccgg  agtgtcgccg  aggatggcac  gggcaatggt
540  gagccgctgt  cgctccccgc  ccgaaagacc  actgttggct  ccgagcacgg  tatcgtagcc
600  gtccggcagc  cgaagcaccc  ggtcgtggat  ttgcgcttcg  cgggccgcga  cctggacctg
660  ttcggcgggg  gcatccggta  ccgccagcgc  gatgttttcg  gcggcggtgc  catgcacaag
720  ctgggcttcc  tgtag
735
<212> Type : DNA
<211> Length : 735
      SequenceName : gi_GDC_MTUB_1515112
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1515112

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgagcgcgg  tattggcttt  gtctgctgcg  gtatcggcac  gccgcgcaaa  ggctgcggag
60   gcccacagcg  cccccagcag  caacggcacg  ccggccagtg  cagccacgcc  gagctgccag
120  gagatcggca  acagggccag  cgcgatcact  gccggcagca  ggatcgcgct  ggtcaacggt
180  gtcaccagat  taaccaccag  gccaacaagt  tccggcccgg  tggccgcgat  cgcctgccgt
240  gccgtcgcgg  tgttttcggc  ggtaaaccaa  tccaaccgga  caaccggaag  ccggtccgcc
300  acatcatgtt  gggtgtggtt  aaggacggcg  aaacccagct  cgataccgat  gcgtgcggtc
360  acggcgtcta  gcacccaccc  ggccacggtc  gcggcggaca  gccagcccag  ccacagccac
420  gcgcgctgcg  gcgcctcccc  gaacaacgcc  gccatcagcg  gcaccagcaa  cacggtgccc
480  accgctcgcg  cgacaacgga  acaaaacgcg  agcagcgcaa  agccgattag  cctggcgcgg
540  tggtcgttcg  gaacaagggc  tatccaggtg  cggatcatcg  ggtgccgtcc  tgcgctgcgg
600  cgaccgccac  ccggctgccc  tggccggtgt  cccacagccg  gcagtagcgt  ccgcccgcgg
660  caagcaactc  ctcgtgggtg  ccgcgttcga  cgatccgacc  atgatcgagc  acgacgatct
720  ggtcggcccg  ggtga
735
<212> Type : DNA
<211> Length : 735
```

95

SequenceName : gi_GDC_MTUB_1515464
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1515464

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgccatcgg tcattcgcga cccagatccc ggtgcagcgc ccgcaccgac agttgctgat
60 cggagcgcag aagtcccatc agtgcttcag cgatcgcgac gctgcgatgc ttaccaccgg
120 tacagccgat ggcgattgtc atatagcgct tcccctctcg gcggtagccg tcgacaacca
180 gggatagcaa ccgatggtag gactcgagga actcagccgc gcccggccgg tgcagcacat
240 agtcgcgcac ggccggatgt tggccggtca gtggccgcaa ctcgtccacc cagtgcgggt
300 tcggcaggaa ccgcacgtcc atga
324
<212> Type : DNA
<211> Length : 324
        SequenceName : gi_GDC_MTUB_1596569
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1596569

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgctacggc ccatacgggc gggccaacct ggccgacatc tggcgccgcc gcgacctgcc
60 acgcgacgcc aaggcaccgg tgctggtaca ggtgcccggc ggcgcctggg tactggggtg
120 gcgccgcccg caggcgtatc cgttgatgag ccatctggct gcgcgcggct gggtatgcgt
180 gtcgctgaac taccgggtgt cgccgcgcca cacctggccc gaccacattg tcgacgtgaa
240 gcgcgcgctg cgtgggtca aggaaaacat cgccgcctac ggcggggatc cgaatttcgt
300 tgccatcagc ggcggttcgg ccggcggcca tctgtgcgcc ctggcggcgt tgacccccaa
360 cgatccgcga tttcagcccg ggttcgaaca ggtcgacacc tcggtggcgg cagcggttcc
420 ggtatacggg cgttacgact ggtttacgac cgatgcgccg gggcgtcggg aattcgtcgg
480 gttgctcgaa acgttcgtgg tgaaacggaa attcagcacg caccgcgaca tcttcgtcga
540 tgcctcaccg atccaccatg tgcgggccga cgccccaccg ttcttcgttc tgcacggccg
600 ccacgactcc ctgatccccg tggccgaagc ccatgcgttc gtcgaggaac tgcgggcggt
660 gtcgaagtcg cccgtcgcct acgcggacct gccccacgcc caacacgcct tcgacgtctt
720 cggctccccg cgggcgcatc acaccgccga ggccgtggcc cgcttcctgt cttgggtgta
780 cgcgaccaac ccgccggcca cgtagtcagc tataggccag ctattgctat ccgcggcac
840 gctccagctc ggccagtgcc ggttcgatgg catcggccat ctcgtcgatg tcgttggcca
900 cctcgggtgt ggtcaccagg ccgaaatcca gataatcctg gtaggagaag caggtga
957
<212> Type : DNA
<211> Length : 957
        SequenceName : gi_GDC_MTUB_1600905
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_1600905

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :

```
atgacggcca gcaggcgctc ggaccacacg gacgcgacgc gtcgagccct cgtcgacgct
60  ggccgttacc tattcgcgcg cgcgactat ggtgacgtct cgatcgaaga catcgtcacc
120 cgtgcccgag tcacccgtgg cgccctggac taccacttcg acagcaagaa agatctgttc
180 cagacggtac tcgaggttgt cgaagccgac ctggtcgccg acgtcgaagc cgccatagcg
240 aaggtcaccg acgcctggat ctgctggtcg tcggcttcca cgccttcctt gacgcggcga
300 ccaaaccgga tgcgctgcag gtcattgcga ttgacgccc gtcagtcgtc gggtggggcg
360 aatggcgccg atcgacatg cgctagggct tggtctgctg gtcggggctc tcgaacgcgg
420 gatggccgcc ggggtgattc agcgcgtacc gttgccacca ctttcgcatc tgctgctggc
480 cgcgctaacc gaatccgcgc tgcagatcgc ggacgcgacg gacaaagacc ggaccagagt
540 cgaggtcgaa cgcgcattta tggccctact cgaaggtcta cgggtgtagc acgcccgcga
600 tccgctacgg caacggacca ccggccgcaa tcgcggccag cgtcgcgaaa tgctccccgt
660 ccagcgacgc cccgccgacc aggccaccat cgacgtcatc ctgggccacg atgtcgccga
720 cgtttttggc gttcaccgag ccgccgtaga gcacccgcac cgtatcggca atcctcggcg
780 aggccaacga ggccaactct ttcggatcg ccgcacacac ctcctgggcg tcggcggcgc
840 tggccacccg cccggtgccg atcgcccaga ccggttcgta ggcgatga
888
<212> Type : DNA
<211> Length : 888
      SequenceName : gi_GDC_MTUB_1616064
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1616064


Sequence
--------

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtcgcgtg ctatccggac aaagccgaaa tcagcatctt cccgggggtag cgcaggctac
60  cgggtatacc tcggccaacg actgggtgtc gctgtattcg cgcagcgaga tgatcatccc
120 gtcacgggtc tcgaagatgc agacgaacgg gctgtcatat cgggtccggt cggcgctcac
180 accgtcgcaa tgcccctcga ccactaccgt ttcaccctcg ttgacgcagc ggatgagttc
240 gatgttgacc tcgaagacct gcttgcgccg ctcgactgct cgccgaaacg tcttcttgtc
300 caattccgta cgggtgacga tgctccagta ggtgaagtcg ttgctgagca gcgcgaagcc
360 ttcgtcgaga tctccgccct cgcagaggct ttgcaggaac atccaggcca gttcggcttg
420 cgggtcgtcg aacggcgtca tcacatcgcc atcttgtctc gggagacagc gtgcggtcaa
480 ttgacgtggt cgtcgaagcg gtggtcacct tcgcggggc ggccggcttc gcgcacacct
540 tggcgccgtt gcgtcgcggt cagcaggatc catgctttcg gtccccggt gacggcacta
600 tctggcggac cagcttgctg cccaccgggc cggtcaccgc gcggatcagc cgtgctgggc
660 gcgacgccgc ccgttgcgtg cgtggggca gcggtgccga ggagtttgtc gacatggcgc
720 ccgccatgct gggcgccgc gacgacgcca gcgatttcgt ccgctgcat ccggccgtgg
780 ccgccgcgca ccgccggctg ccgaacttgc gcctgggccg caccggccag gtgctggaag
840 ccttga
846
<212> Type : DNA
<211> Length : 846
      SequenceName : gi_GDC_MTUB_167239
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_167239


Sequence
--------

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcggttac gctcggaaag cgcgggcctc gcccacgcgg cggatgatgt cagcggggtg
60  gtcctcggcg acgacccgga ccacgatcca cccgtagcgg tgctggactt tctcgtgccg
120 gaggatgtct ttccggtagt ggtagcgact ggtcagatgg tggtcgccgt catactcggc
```

```
180 cgcgaccttg atgtcttgcc agcccatatc caaatgggct tccgcccagc cccattcgtt
240 gcgcaccgcg atctgcgtct gggggcgcgg aaagccggcg cggatcaaca acaagcgcag
300 ccaggtttcc ttgggggact gggcaccgcc gtcgacgagg tccagagcgg ctcttgcggc
360 cttcatgcca cggcggcccc gatagcgctc gatcagcggc tcgacgtcgg ccaccttcaa
420 atcggtggcc tgtatcaggg cgtcgacggc cgcgacggcg gggtccaatg gaaatcgact
480 ggtcaggtcg agcgccgttc gctccggtgt ggtcacgcgc atgccctcga tgacgcagat
540 ctcgtcgggc tcgatgcgct cttcccagac ttgcagcccc ggggcacggc ggcggttggt
600 gtcgatgatc gcggcgggaa gatccgcgtc gatccacttg gcgccatgga aggcagaagc
660 cgagtagccg gccagcacgc cgcggcggcg cgagcgcagc cacagcgctt ttgcacgcaa
720 ttgcgcggtc agttccacac cctgcggcac gtacacgtct ttatgtag
768
<212> Type : DNA
<211> Length : 768
      SequenceName : gi_GDC_MTUB_1672449
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1672449


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgggtgtgc gcgccgccgt cggcgtagat gatgtcaccc gtggtcgccg ccagccagtc
60 agacagcagc gcgcacaccg tcttggcgac cggcgtcgca tccttcatgt tccagccgat
120 cggagcgcgc tgatcccagc cctcctcgag cagctggatc tgggcgccgg cctcctcgcc
180 gagcgcaccg ccgacgatcg cactcatcgc cagcgtccgg atagggcctg cggcaacgag
240 attcgaacgc acaccgtact tgccggcctc gcgcgccacg aacctgttga ccgactccaa
300 cgcgctcttg gcgaccgtca tccagttgta ggccggcatc gcccggctcg ggtcgaagtc
360 catgccgacg atggaacctc cggggttcat gatcggcagc agcgccttgg ccatcgaagc
420 atacgaatac gccgagatgt ggatgccctt ggacacatcc gcgtagggcg cgtcgaagaa
480 cgggttgatg cccatcccgg tctgcggcat gaacccaatc gaatgcacca ccccgtcgag
540 cttgttgccc gccccgatcg cctcggtcac ccggccggcc aagctggcca ggtgctcctc
600 gttttgcacg tcgagttcga gcagcggggc ctttgccggc agccggtcgg tgatgcgctg
660 aatcagccgc agccggtcga acccggtgag caccagctgg cgccctgct cctgggctac
720 ccgtgcgatg tgaaacgcga tcgacgagtc ggtgatgatt ccgctaacca gaatccgttt
780 gccgtccagc agtcctgtca tgtgcgtcct tgtgttgtgt cagtggccca tacccatgcc
840 gccgtcgacc gggatgaccg caccggagat atagctcgca tcctcggaag ccaggaagct
900 gaccaccccg cgcacctcgg cggggtgcc gacccgcttc gctgggataa attgcagcgc
960 ccctgctga atccgctcat ccagcgcgcg ggtcatatcg gtgtcgatgt agcccggggc
1020 caccacattc gcggtcacgt ttgccttcga cagctcgcgg cgatcgagc gggccatgcc
1080 aatcactccg gccttggagg ctgcgtagtt ggcctggttg ccgatgcccc agctgccgga
1140 gaccgaacct atgaatatca ttcgaccgaa tttgttgcgc tgcatgctgc gcgatgcccg
1200 ttgagccacc cggaacgccc cggtgaggtt ggcgttgatg accttctcga actttcctc
1260 ggtcatccgc atgaggaatg cgtccgcgga tag
1293
<212> Type : DNA
<211> Length : 1293
      SequenceName : gi_GDC_MTUB_1673708
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1673708


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atggtgccga gcatgagggt gcgctcggat tgggagccga tcgcccagag ccgctcccgg
```

98

```
 60 ctcgcggtca cggcaccgcg caacacctcc gggggtcgct tcatctggat tctcctcggt
120 tctgcgcgaa acggtagcag agcgccatgg ttgccaacgc ggtcgccggg cagtctagac
180 cggatcttcc tcgtggcaac cgacaacagg acgtcgttgc cgaaagggcg ctgggcaccg
240 acatctagga tgaacccaca gccacgcccc gacgttatgc catggcgaag agcgaccggc
300 aggagcggga acccagtgaa gcgagcggctc atcaccggaa tcacaggacc ggacggctcg
360 tatctcgcta agctcccgct gaagggatat gtggccgctg gtagcccggc cgaggtctat
420 ttctgctggg cgacacggaa ttatcgcgaa ttgtatgggt tgctcgcggt caacagcatc
480 tggttcaatc acgaatcacc gcgtcacggc gagacattca tgactcgtaa tcctgcacca
540 tatcgcggtc ggcaacgagg cgctgatcga tgcgcagacg ctgatgcgcc ggcccacccg
600 gataggtatc agtattgggg cgttccggcc agcgtacgag gcgtgatcga ccgcgcaatg
660 ggtgtttgcg ttgagtaa
678
<212> Type : DNA
<211> Length : 678
     SequenceName : gi_GDC_MTUB_1699549
     SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1699549


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgagcggtc agccatcggc tttgcgccga cctacggtgt ccccgtcggc gtgtcgccga
 60 cctacggtgt cgaagtcaaa gccaaagatc gacaggatga ccagcaggat ggcgccaccg
120 actaccgacg gatcggcgac attgaacacc ggccaccagc cgaccgacaa gaaatcgacg
180 acgtgcccgc gcagcggccc cggtgcccga aagaagcgat caaccaggtt gcccatggca
240 ccgcccagga tcatcccaag acccagcgcc caccacggcg ataccagccg ccgccccatc
300 cagaaaattc cgaccacgac acccgtcgca atcagcgtca aaacccaggt gtatccggtc
360 gccatcgaga aggccgcccc agaattacgc accagagtcc aggtcaccgt gtcgccgata
420 atcgacaccg gctggccggg cggcaacagt tggacagcta ccaccttggt gacaatgtcg
480 agtgtgagca ccaccacagc gaccgacagc agcatgcgca gccgtcgcgg cggcgcggga
540 gcgttaggtt cccccgcccc cccggcttcc tcggtcgagg tcagcggatc agccgatcct
600 gttggttcgt caggcacacc atcatcatcc cctagggccg atatggcccg cccagacccc
660 gcggccggat gggagcaaac cacgtgcgca atgatcccat catggcccgc ctcaccgtca
720 tcactactgg agggacaatc tcgaccaccg ccggccccga tggggtgcta cggccaaccc
780 attgcggggc gacgctga
798
<212> Type : DNA
<211> Length : 798
     SequenceName : gi_GDC_MTUB_1742061
     SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1742061


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcccccga ataggccgga acgccggtta gggaaacctc taacagcgcc gcttcgacgc
 60 gcaccagcac atccccttcg cgacggtccc ggatcggtcg gaaacccacc gaaaacgagt
120 cgacgacacc agcttttacg ttcgccaaag cctcgtcgcc gtccggggtg tccgcaatct
180 cgaacgcccc gaacaagccg tgaggctcct cccgcaactc aacggcccgg cccaccgggt
240 agcgggttcg agcgtcgtga gagaccagca gcttcaattt gtggccgcgc tcggcgatgg
300 agcgccgaaa agcgccagga gcgaacattt cctggaactc gccgtcgaag tcgcggacgg
360 tggtcgcctc gttgtagggc acgatggtgc cgtgcacggt tcggccttcg ccagaccgca
420 gctcggccat gcggaaaagg atgctactca aaattcggcc accacctagc agacgcaaga
```

99 gg

```
480 aacgcgcgga atcgcttgtg gcgcatggcg gccgctatcc gggttccagc cgccccgcgg
540 cgactgcccg gcgtcagcgg atgccgagat gccaaactcg attgtatcac acacaaaagg
600 tcatcaccgg tccgggggcaa acgggttgag cccgtcgccg tcgtcgcccg gcgccaccgc
660 cagtcgctgc tcggcggccg gggtcaggcc aaactcggag gccaagcgca gcagatgcat.
720 gcgcgccgtc tccgcaaccg tcaccgccgg gttccggtgc acgacaccgg atttcggtga
780
<212> Type : DNA
<211> Length : 780
      SequenceName : gi_GDC_MTUB_1782153
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_1782153

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgtggaaat ggaagccgcg cttggcattc caccgggcaa cctggcggcg acgctggacc
60 gctacaacgc ctacgccgcg cgcggcgcag atcccgattt ccacaagcag ccggaattcc
120 ttgcagcaca agacaacggg ccgtgggggg cgttcgacat gtcgctgggc aaggcgatgt
180 atgccggatt cactctgggc gggctggcca cgtcggtgga cggtcaagta ctgcgcgacg
240 acggcgcggt ggtggccggc ctgtacgcgg tcggggcatg cgcgtccaat atcgcccagg
300 acggcaaggg atatgccagc gggacccagc tgggtgaggg gtcgtttttc gggcgtcgcg
360 ccggagcgca tgcggcagcc cgagcgcagg gcatgtaagc ctcctcgcgc cgcgactggg
420 aatcctgcga cgcgacacgc cgacaaggcg tcgtga
456
<212> Type : DNA
<211> Length : 456
      SequenceName : gi_GDC_MTUB_2060659
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2060659

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgtggcccc gtatttccgc ggcgccgtcg aatcggcgat cgacagttgg cggcgtgtgg
60 tgtcgacggc ggcccaactg ggtatcccga ccccgggatt ctcgtcggcc ctgtcgtatt
120 acgacgcgct gcgcaccgcg cggctgcccg ctgcactcac ccaggcccag cgcgacttct
180 tcggcgcaca cacctacggc cggatcgacg aaccaggcaa gttccacaca ctatggagtt
240 cagaccgcac cgaagtaccg gtgtagcggg ctagaactaa aaggggggtaa aggggtaagt
300 gatgagattt ctagacgggc acccacccgg gtacgacctg acatacaacg acgtgttcat
360 cgttccgaac cgatccgagg tcgcgtcgcg cttcgacgtc gatttgtcca ccgccgacgg
420 ctcgggcacc accattccgg tagtggtcgc caatatgacc gcggtagccg ggcggcggat
480 ggccgagacg gtcgcccgcc gcggtggcat cgtaatcctg ccgcaggatc tgccgatccc
540 ggcggtaaag cagacggtgg cgttcgtcaa aagccgggac ctggtgctcg acacccccagt
600 gacgctggca cccgacgatt cggtgtccga cgccatggcg ctcatccaca agcgcgcaca
660 tggcgtcgcg gtggtcatcc tcgagggtcg cccgatcgga ttggtgcgcg aatcgtcctg
720 cctgggcgtg gatcgcttca cccgggtgcg cgatatcgcc gtgacggact atgtgaccgc
780 tccagcggga accgagccac gcaagatctt cgacctgctg gagcacgccc cggtcgacgt
840 tgcggtgctg accgacgccg acggcacgtt ggcgggagtg ctaagccgca ccggggctat
900 ccgcgccggt atctacaccc cggccaccga tag
933
<212> Type : DNA
<211> Length : 933
      SequenceName : gi_GDC_MTUB_2093062
```

100

SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2093062

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgggtatat ctcccggcga tcgcggggat cgtgttcgtg gcaatgccgc tggtcgcgat
60 cgccatccgg gtcgattggc cgcgtttctg ggcgctgatc actactccgt cttctcaaac
120 ggccctgctg ttgagcgtga agaccgccgc ggccagcacg gtgctgtgcg tactgctggg
180 cgtcccgatg gcgctggtgc tggcccgcag ccgcggacga ctggtgcggt cgttacgacc
240 gctgatcctg ttaccgctgg tgctgccgcc ggtagtcggg ggtatcgcgt tgctctacgc
300 gttcggccgg ctcggcctga tcgggcgcta cctggaggcg gccggcatca gcatcgcatt
360 cagtaccgcg gctgtggtgc tggcgcagac ctttgtctcg ctgccgtatc tggtgatttc
420 cctagagggt gcagcccgca ccgccggagc cgactacgag gtggtggcgg cgacacttgg
480 ggcgcggccc ggcactgtct ggtggcgcgt gaccctgccg ttgctgctcc cgggcgtggt
540 gtccggatca gtactggcgt ttgcccgctc gctcggagag tttggcgcga ccctaacctt
600 tgccggttcc cggcaagggg tcacccgtac ccttccgctg gagatttacc tgcagcgggt
660 gaccgatccg gacgcggcgg tggcattgtc actgctgctc gttgtggtag cggcactggt
720 ggtgctgggt gtgggtgctc gtacgccgat cgggaccgat accaggtagc cggtcatgag
780 caagctgcag ctgcggccga tcgtcgccga ccggcgtttg gacgtcgaat tctcggtgtc
840 cgcgggcgag gtgcttgcag tgctcgggcc caacggtgcg ggcaagtcca ccgccctgca
900 tgttatcgcg gggctgcttc gccccgacgc gggcttggta cgtttggggg accgggtgtt
960 gaccgacacc gaggccgggg tgaatgtggc gacccacgac cgtcgagtcg ggctgctgtt
1020 gcaagacccg ttgttgtttc cacacctgag cgtggccaaa aacgtggcct tcggaccaca
1080 atgccgtcgc gggatgtttg ggtccgggcg cgctag
1116
<212> Type : DNA
<211> Length : 1116
        SequenceName : gi_GDC_MTUB_2105797
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2105797

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcccacgc cggtcccagc ccgaactggg acgccgtcgc gcagtgcgaa tccggggggca
60 actgggcggc caacaccgga aacggcaaat acggcggact gcagttcaag ccggccacct
120 gggccgcatt cggcggtgtc ggcaacccag cagctgcctc tcgggaacaa caaatcgcag
180 ttgccaatcg ggttctcgcc gaacagggat tggacgcgtg gccgacgtgc ggcgccgcct
240 ctggccttcc gatcgcactg tggtcgaaac ccgcgcaggg catcaagcaa atcatcaacg
300 agatcatttg ggcaggcatt caggcaagta ttccgcgctg acggttggcg gcgtgtgcgg
360 tctatgacca ggtcgacgta tgtgtttgga tcaggtcatg gaaggttcgg ccacagttca
420 catggcagcg ccgccggaca agatctggac attgatcgcg gatgtccgca ataccggccg
480 gttctcgccg gaaaccttcg aggccgagtg gcttga
516
<212> Type : DNA
<211> Length : 516
        SequenceName : gi_GDC_MTUB_2133554
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2133554

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcgaccgg gccaccgcca ggtcgatgga tgccgccgtg gccaaccgtt gtgcggtgct
60 catgaacgcg tcggcctcgt gcgggttgtc ggtgccttcg gcctggcgca gcagggctgc
120 gatgcgggcc agcatcttgt cgttggtcat ggcgccaaaa ctagtggagg gctgcgacag
180 gtcggctcgg cctacaaccg ctcggtgagc caggcgacca catcgtcgag cacctggttg
240 cgctccggct cgttgaacac ctcgtggtac agcccgggat actccttcag ctgcacgtcg
300 gccgatccca cacattcgac caggcgacgg ctgccctcga tggggatcag ccggtcatcg
360 gtgccgtgca gcactagcag cggcgcggtc aatgccggtg ctcgccgcgg catggtctcg
420 cccacctgca gcagcgcgcg gccaatcccg gccggaaccc gtccgtggtg cacgagtggg
480 tcggtgttgt aa
492
<212> Type : DNA
<211> Length : 492
      SequenceName : gi_GDC_MTUB_214625
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_214625

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcgccggc tccgctcttc agatccacgg tgccatcgcc ttcacgtggg agcacgacct
60 gcacctgtat taccgccggg ccaagaccac cgaggcgctt ttcgggagca gcgctcgaaa
120 tcgtgcgctg ctcgccgaac gcgcgggggct tgtgaaagcc taggcgccca gcgcggccag
180 cgccgcttcg tagttgggtt cttgcgcgat ttccggcacc aattccgtgt aggcgacgtt
240 gccgtccgcg ccgatcacca cgattgcgcg ggcgagcagc ccggccatcg gcccgtcggc
300 gatggtcacg ccgtaatcct cgccgaagct gtcccggaat gccgacgcgg gcatgacgtt
360 ttcggtgccc tcggcgccgc agaagcgctt ctgggcgaac ggcagatcct tcgagacaca
420 cagcacggta gcgccacttg ccgccgcacg ctcgtcgaag gttcgcacac tcgtcgcgca
480 caccggtgtg tccacggatg aaagatgtt cagcaacacg gacttacccc ggaactggtc
540 gctgctgatc accccccagat cgcccccggt cagggtgaag gccggggccg gggatccgac
600 agcaggtag
609
<212> Type : DNA
<211> Length : 609
      SequenceName : gi_GDC_MTUB_2183418
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2183418

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcgcgggt ccgggcggac gcagatacaa gaccacgccg ctgccctgag ccgacatcct
60 cgccagcgcg ccgttgagtt cctcgccgca gcggcacgcc gtcgagccga acacgtcgcc
120 cgtcaggcac tcgatgtgga cgtgcagcgg cacgggcacc ccggcaccga ccgcacccac
180 gatgaccgcc aaatgctcgc cgaggtcgta aacgtcacga aagccgatga cacgcgaggc
240 gccggcccag gtgggcagcg tcgctgccgt aaaccggacc acctggggct cgatccgccg
300 gcgatacgcc accagctccc cgatcgagac catggccagt ccgtgttcga cggcgaattc
360 gaccgactcg gcgtggtgcg ccatctggac gggattatcg gcgagacga tctcgcagag
420 cgcggcggcc ggccgccgtt ccgccaggcg gccaggtcg acggccgcct cggcgggtcc

```
480 ccgccgaccc agcacaccgt cggcttgcgc ctgcacgggc accacatggc ccggacgttg
540 gaaatcggcg gcgacggagg tggccgaagc cagtgccgcg atggtccagg cgcgatcgct
600 cgccgagatt ccggtgccgg tgccgcgaac gtcgaccgac acgcaatgcg tggtgtctcg
660 gtcacacatg ggcggcaggt gcagtcgctc gcattcggcg cccggcagcg cgacgcgcaa
720 ataacccgag gtgtgccgga ccgcaaaggc aaccagccgc ggcgtcgcgg cctgggcggc
780 gaagacgaga tagccatcgc cattggggtc gccggtcagg accacggcgt gaccgcccgc
840 catcgccgtg atcgcacgac gtacccgcac atcggtcgtc ttcatcgaga ctccaaccgg
900 cggaaccggc taccgtga
918
<212> Type : DNA
<211> Length : 918
      SequenceName : gi_GDC_MTUB_2192571
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_2192571

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgaagacag ctatttctct gccggatgag acgttcgatc gggtatcgcg gcgtgcgagt
60 gagctcggca tgagtcggtc cgagttcttc acgaaggctg cgcagcgcta cctgcacgag
120 ctggacgccc aattgctcac gggccagatc gacagggctc tagagagcat ccatggcacc
180 gacgaagcgg aggccctcgc cgtggccaac gcataccgcg tgctagaaac catggacgat
240 gagtggtga
249
<212> Type : DNA
<211> Length : 249
      SequenceName : gi_GDC_MTUB_2234641
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_2234641

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtctacat ccacgacgat tagggtttca acccagactc gggatcgtct ggccgcccaa
60 gcccgcgaac ggggaatctc gatgtcggct ctgctcaccg aactggccgc ccaggccgag
120 cgccaggcaa tcttccgcgc cgaacgcgag gcctcgcacg ccgagacgac cacccaggca
180 gtccgcgacg aggaccgcga gtgggagggc acggtaggcg acggccttgg ctga
234
<212> Type : DNA
<211> Length : 234
      SequenceName : gi_GDC_MTUB_2320829
      SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_2320829

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggcgacca gcacctcgcc ggccggtggg ctgccgcagg cccgctcgca gccgacgaaa
60 tgccgatgcc cggctgactc cacgttcagt gaccgcgcgg cgtcggcccg tacgtcggcg
```

```
120 gccgagtgcg cgcagccggg gctgccggtg caggcgctga tgttcagcca gggggagttc
180 tcgtcgaaca ccaggcccag cggcgccagc acccgcagcg cggcgtcggc cgtcgcgtcg
240 tcgaggtcgc agatcagcac cgatcgccac ggcgtgatca ccagcggggc ctcgatcgcg
300 gccaggcatt ccgcgacccg ggcgggcaag accccagcg gcaccgcggc gcccagcgtt
360 acccggctgt catcctgggg tatccagccg acgggcgttt tggtgacggg ccgaacggat
420 gggcccagct cgacaccgga ctgcagctcg ccgatatcgg ctaattccgt tactcgccag
480 gcggtttcgc ggatcttgac gaaacgcaac gcgaccctga tcagggtctc ggcgacatcg
540 gccacccgca cgccggtgtc acgtccggtc aacagcagtc ggggaccgtc ggggaacacc
600 tgcacgccga cgtcggcacc caggccggac acgtcggcgc ggcgtcgtc gagaccgaac
660 cagaaccggc cgcccagttc cgccagccgg ggctcggcgc ggatcgccgc gtcgagctca
720 ccgacccatg cccgcacgtc ggctagcccg ccggcccggc cggacagcgg cgaggcgacg
780 atattgcgca cccgctcgtg tgttgccgac ggcagcagcc cggctttggc gaccgcgtcc
840 gcgaccgctg ccacgtcgcg gatcccgcgc aactggacat tgccgcgcgc ggtcagttcc
900 agtgtcgcgg agccgaagtc gctggcgacg ctggccagcg tcgccagttg tgccgcggtg
960 atcatcccgc cgggcagccg gatccgcgcc agcgccccgt cggcggcctg gtgcggccgc
1020 aacgcaccgg ggcaggcgtc cgcgtcacgg gtcccggcca cccgtccacc gtacgggaga
1080 atgggtcgcc gcctcgccgc gctcaggtcc cgtcgggagg ccgaggatca gggtcagggc
1140 gttttcgatt gcgcgcatcg tggcgggttt gagggcgccg agtcgctcca cgagtcgggc
1200 acgagcgacc gcgcggacgc cgcggcacac cgcgacgcta tcggcagcta caccttctga
1260
<212> Type : DNA
<211> Length : 1260
      SequenceName : gi_GDC_MTUB_2321250
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2321250

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgacgggcc gtgtccgaca gaccggcata acccgtctcg tcgtacatca gcggggcccc
60 gtccttccac agcgactgat gacagtgcat gccggacccg ttgtcgccga acagcggctt
120 gggcatgaac gtgaccgttt tgccgttctg ccaggcggtg ttcttgatga tgtacttgta
180 caactgcatg tcgtcggcgg cgtgcagcag cgaattgaac tggtagttga tctcggcctg
240 tccgccgctg cccacctcgt ggtggccctt ctccaggatg aagccggagt tgatcaggtt
300 ggtcagcatc ttgtcgcgca ggtcgacgta ttggtcgttg ggggccactg ggaaataccc
360 gcccttgtgg cggaccttgt agccccggtt gggactgccg tcggcctcgg tcgccgcgcc
420 ggtgttccac caccccgaga tggcgtccac ctcgtagaag gagccgttgg cgcgcgagtc
480 gaagctcacc gaatcgaaaa tgtagaactc ggcctcggcg ccgaagtatg cggtgtcggc
540 gatgccagtg ctgatcaggt agttctcggc cttgcgggcg atgttgcgcg ggtcgcggga
600 gtacggctcc agggtgaacg ggtcgtgcac aaagaagttg atattcagcg tcttggccgc
660 gcggaacggg tcgatgcgcc ccgtctcggg atcggaagga agcaacatgt cggattcgtg
720 gatcgactgg aacccgcgaa tcgacgagcc gtcaaaggcc aagccgtcgt caaacacgct
780 cttgtcaaag gccgaagccg gaatcgtgaa gtgctgcatg atgccaggca ggtcacagaa
840 ccggacgtcg acatattcga ccttctcgtc cttggcaagt ttgaagacgt cgtcgggcgt
900 cttttccgtc acagaatgct cctttactgt atccgcggcc gacgctatgg agccgatatt
960 gcccgtcagt caaccccgtg ttgcgcagac gttactgacc gtgccgccca ccactga
1017
<212> Type : DNA
<211> Length : 1017
      SequenceName : gi_GDC_MTUB_2487508
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2487508

Sequence
```

104

```
---------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggcgggcg tttgcgcgct attctccggt gcttcccgct ggccgtctgg tgaacttcgg
60 caccgtccac agggttcccg ccggggtccg agccggctac gatgcacctt tccccgacaa
120 aacgtatcaa gccggcgccc gggcgttccc acggttggtg ccgacctcac ccgacgatcc
180 ggcggtaccg gccaaccgcg cggcatggga agccctgggc cggtggggaca aaccgttcct
240 tgccatcttc ggttatcgcg acccgatact cgggcaagcg gacggtccgc tgatcaagca
300 cattcccggc gcggcgggtc agccgcacgc ccgcatcaag gccagccact tcatccagga
360 ggacagcgga accgaactcg ccgaacgcat gctctcctgg cagcaggcaa cgtaaccgcg
420 acggctgcgg acgaaggatc ggcagaatgg cgatggagat ggcgatga
468
<212> Type : DNA
<211> Length : 468
     SequenceName : gi_GDC_MTUB_2567990
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2567990

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgaccgaca acgagtgccc ggccgacagc cgacggcgcc atgtcctgcg gctcgccctg
60 ttcgccggga ttttgctggg gctgttctac ctggttgcgg tggcacgagt catccacgtc
120 gacggggtcc gtagcgcgat cgtggtggcg acgggtccga tcgcacccct ggcgtacgtt
180 gtggtgtcgg ccgcactcgg cgcgttgttc gtcccgggcc cgatcctcgc cgccggcagc
240 ggggtgctgt cgggccgct actagacacc tttgtgaccc tgccagcttt ctcggccggc
300 gcgcaggccg gaatgacgcc caggcgctgc tgggtgtcga tcgcgcccat cgcctcgatg
360 cacagatcga acggcgcgga ttgtgggcgg tggtcggtca gcgcttcgtc cccggcatct
420 cggatgcgct ggcctcgtac accttcgggg cgttcggagt tccgttgtgg cagatggtcg
480 ttgggtcgtt catcgggtcg cgccacggg tgttcgtcta caccgcgctg ggcgcgtcga
540 tcaccaacct gtcgtcgccg ctggtttact cggcgatcgc ggtgtggtgc gtga
594
<212> Type : DNA
<211> Length : 594
     SequenceName : gi_GDC_MTUB_2577106
     SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2577106

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgtgggcgg tggtcggtca gcgcttcgtc cccggcatct cggatgcgct ggcctcgtac
60 accttcgggg cgttcggagt tccgttgtgg cagatggtcg ttgggtcgtt catcgggtcg
120 gcgccacggg tgttcgtcta caccgcgctg ggcgcgtcga tcaccaacct gtcgtcgccg
180 ctggtttact cggcgatcgc ggtgtggtgc gtgaccgcca tcatcggggc gttcgccgcg
240 cggcgttggt accggaagtg gcgtgcgcgc ccgcgccggc ggtgcggcct ggctcagctc
300 acgaccggta gtcagcaacg ccacacgagt caccggacac cggcgggcgt cgtcatgccc
360 ggttcactgt ccgagcaccg ccgtctccgt caagaagcgc cggatcgcat cgagcatcac
420 ccgcccatcg agtag
435
<212> Type : DNA
<211> Length : 435
     SequenceName : gi_GDC_MTUB_2577486
```

105

SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2577486

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtatatac gtttttatcg cgattctctt gcagagcccg ccacagacat atacgctttt
60 gcctatgttt cgttcaacaa ggaggccggc acatggcaca cccctgcgca accgacccgg
120 aactatggtt cgggtacccc gatgacgacg gcagcgacgg cgccgctaag gcacgcgcct
180 atgagcggtc ggccacccaa gcgcggatcc aatgcctgcg ccggtgcccg ctcctacagc
240 agcgccggtg tgctcaacac gcggtcgagc atcgggtgga gtacggcgta tgggccggca
300 tcaagcttcc cggcggccag taccgaaagc gcgaacagct cgcggcagcc cacgacgtgc
360 tgcgtcggat tgccggcggc gagatcaatt ccaggcagct cccggacaat gcggctctgc
420 tggcccgcaa cgaaggactc gaggtcaccc cggtgcccgg ggtcgtggtg cacctgccga
480 tcgcacaggt tggcccacaa ccggccgctt gatgcccggt cggcaagccc ggcagttgcc
540 aaacccagcg tgatcaggct cggctcgcga gttcggcgaa gaagtggctc gcctgatcac
600 ctaccatcgg ccaggatctg cgtgtcatca cgacgctcgc caaggaggtt gttgtggtgc
660 tatcgacggc ctttagccag atgttcggaa tcgactatcc gatag
705
<212> Type : DNA
<211> Length : 705
        SequenceName : gi_GDC_MTUB_26830
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_26830

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgtctgcgg ttttaccggc tcggtgcatt cgcgcgctag ccgatagggt ctatcgccat
60 gtccggtgcc acggtgggtg cgcgcgaaat caccatccgc ggagtcgtcc tgggcgcatt
120 gattaccttg gtgttcaccg cggccaacgt gtacctgggg ctaagggttg gattgacatt
180 cgccacttcc ataccggccg cggtgatctc gatgggcgtg ctgcggttgt cgccaaccca
240 ctcagtggtg gagaacaata ttgttcagac gatcgcgtcg gcggccggca cgctgtcgtc
300 gatcatcttc gtgttaccgg cactgctcat gatcggctgg tggagcgggt ttccgtactg
360 gacaacggcg gcggtgtgtg cactgggcgg gatccttggc gtcatgtact caattccgtt
420 gcgccgcgca ctcgtcaccg gatcagacct gccgtaccca gaaggcgttg ccggagccga
480 ggttctcaag atcggtga
498
<212> Type : DNA
<211> Length : 498
        SequenceName : gi_GDC_MTUB_2690012
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2690012

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgggcccga tgaacgggtt cctgagttgg tgggacggcg tcgagctgtg gctgtccgga
60 ctcccgttcg cgctgcaggc gttggcagtc atgccggtcg tgctggcttt ggcctatttc

106

```
120 accgcggcat tgctggatgc cctgctcggc cgggtcattc agttgattcg ccgcgcccgc
180 cgccccgatc aggcgcccag gtag
204
<212> Type : DNA
<211> Length : 204
        SequenceName : gi_GDC_MTUB_2698040
        SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_2698040

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atggcggacg atgtgagcgg cgcggtgtac cgggccggca cggcccacgg tcggccgacc
60 ggtcgcattg aacaccgcga ccgtcaggtc gtgacgcgcc gggcgactga tacgcgcgcg
120 gaactggacg ggctgtccga ccatcagctc gccgaagtcc agcgctcgcg cgaaaaccac
180 tacccggccg gatgtctcgt catcccgcag ccgttgaacc gtcgcccgga acatcaaccg
240 gccccgcccc agcgacactg ggctctcgct gggggtgacc gtgaccagcg cggaggtgcc
300 aaatgccacg gtgattgggt ggcgatcgac cgcctcggag cgcaacgcga ccgcaagccc
360 gtaccccgcg cccaccatac cgaccgcgac caggccggcg ctgatcgaac ccagtcgcgg
420 agcgtgccac gaccggccgcg ccacacacca ccacagtgcg ccgccgccga gggccaccac
480 gacgcagcac aaggcacaca cgttgccgat cggccacacg atcccggccg ccgtcacaat
540 ccagctgacc agcgccgccg ggaccaggcg tacgtccaaa cgggacgcgc cgaagcccat
600 atggcgcacc ggtatcagac acggaccaga ttgcgccgct tgtccagccg cgccggaccg
660 atgccgtcga cgtcggcaag ctggtcgacg ctggtgaacc taccattgcg ctgccgccac
720 gccacaatcg ctgcggcggt gaccggcccg atgccgggca gggcgtccag ctgctccacg
780 gtcgcagtgt tgaggtcgag cacctcagct gtcttaggag ctgtcttagg gcctgtcgtg
840 gctgtgcccg aggtacccgc cggtcccggc gtccccgcac cgaccgagct gcccagcacc
900 ctcggctgtc cgagggcgg agctagcccg accacgatct gctcaccgtc accaagctgc
960 cgagccatgt tcagtccgac ggtgtccgcg ccgtctaccg ctccgccggc ggctgtagc
1020 gcatcggcga tccgcgcgcc cggcgccagg gtgacgagtc ctggggtgtg caccaggcca
1080 accacgctga ccaccaccgg caggccggaa cggtccggcg agcccgggct tgccgacgac
1140 ctagggttcg tcggcgaaac cggctctacc ggaggaagtt tggctgacat taccggctca
1200 gtccggtcgc ggatcaaggt gaataccgtc accagcaccg cgagggcggc gatcaccgcc
1260 aatgcgacgg cgccggcacg gcccggatct gcgcgtatcc tgtccgccca accttgccca
1320 cgggaagtgt cgggaagcca cgcggcagc agcgagttcg gatcgtcgcg tggctcgtcg
1380 tggtctggac cgtcgtccgt tggatcgtgt ggctccgggt ctaagtgtgc agatgcggcg
1440 tgcgagtcga tatccgggac ggcaccgagc cgcctttgca gtcgctcggc gggcagttct
1500 gttcgcatgg gccgaccgca gctgcgggga ccgccagaac cggcgcgcac gacggcgtcg
1560 cgctgccctg ctgtggatca atccgaggct gtggacaagc cgctttggcg atggatcaag
1620 atgggacaaa ccgcgccaac atccccgaac aaccagcacc gggctgcgac gtccatccgg
1680 actcggctca ccgcgatcga gagtgtactc ggcaacgcga tccgcgagtg ctga
1734
<212> Type : DNA
<211> Length : 1734
        SequenceName : gi_GDC_MTUB_2712275
        SequenceDescription :
```

Custom Codon

Sequence Name : gi_GDC_MTUB_2712275

Sequence

```
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgaggggca ctgcctacgc gaccagacgc tcgatgctgc ccaacacccg ggcggtgtgg
60 ctggccaccg tcgtgcagtg cgtgaccggc gggctggggg tgacactgat ccgcagacc
```

107

```
120 gcggccgccg tcgagaccac gcgaagccgg ctggaactcg cccgattcgt cgcccctgcc
180 cggcgcgacg aatcggtttg gtgtttagct ctttcggcgg ccgcgagaag tcctaccagc
240 gtcttgccgg gattatcggc aagctga
267
<212> Type : DNA
<211> Length : 267
      SequenceName : gi_GDC_MTUB_2725593
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2725593

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcgcagag tattcagcgg ttggacaacg ttggtccgct gcagcaccgc agcgaccacc
60 gtcacgatca gggcgatgac aaagcacgtc ccggtaatcc actccagcga accgacccgg
120 ccgctgacgc cgcgaaagcc ggtggatccg gtgcgtcggt gctgcagcca actgcgtcag
180 ccgaatccga ccacactgaa aaccgcgaag agtgccagcg ctaagtcggc cgcggtggtc
240 gttcgcatca gcgggtctcc ttcggtgcgt agcagtggtc atgaaccgtt gtggcggttg
300 gctcgcaggg ccgcatcgat cgcggcggcg gccggtgcgc agtcgccgac accggacacc
360 aaagttgcca gcgcacccgc agcgcaggcc cgccgcaatg cgcgcagtcg ctcggccggc
420 gaacctgggt tgccggccga attcgcagca aggaccccgg caaatacgtc gccggcgccg
480 gcggtatcca ctggcgttac cgttggggcg ggtacctcga acacccgtc cgcgccgacg
540 taccgggcac cgcgcacacc cagggtgatc acgaaatgtg ttggtggcga cggccagtcg
600 tttgcctcat gctcgttggc gatcaccacg tcggcgatag cggccaagtc ctgcaaggag
660 cttcgatcct ggccggctgg ggaggcgttg accatgacaa ccgcatcggc cgactgggct
720 gcccgcgcgg ctgccagcgc ggttgcaaca ggaatctcca actgggtcaa cagtacatcg
780 cagttggcga cggccgaggg taccggagtc agatgtgcat tggcacccgg cgccaccagc
840 acggtgttct cggcgctggc atcgaccacg ataatcgccg tcccgctcgg tccgggcacc
900 gtgacggtcc tgtccagtcc aacggcgttg gcgcgcaggt gggcccgcag ctgggcggcg
960 gctggatcgt cgccgaatgc accggagaac tgtacctgcg cgcctgcgcg cgctgcggcc
1020 accgcctggt tggcgccctt cccgcctggc gttcgggtca acgacgccgc aagcaccgtc
1080 tcgccggggc gcggaagcgc gtccaccacg aacgtcaggt ccatgttcac gctgcctacc
1140 acgcacaccc ggggcgccat ggggccgacg ttagtctcac tggcgtttgc catgcttgct
1200 gttggctga
1209
<212> Type : DNA
<211> Length : 1209
      SequenceName : gi_GDC_MTUB_2733212
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_2733212

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgagcgctt ctgcgtcagc cgacaaggtc gtatgcgagt gctgcgagct ctgtgttcct
60 aaacagctcg cgtcagcgat cgcaaccca tacggactcg tccgtgggtg cgctgtcgc
120 atctgtaacg agcaccaagg ccagccggtc aagatggcgc aagaccacga agaggaggtc
180 cgcatccgtt ggggcgagac ggtggacgaa ctccacgctg cgctggaccg cgccgggcca
240 aggccaggga cgtggtgtac gagtgaaggt tcctcgcgtg atccttcggg tggcagtcta
300 ggtggtcagt ctggggtgt tggtggtttg ctgcttggcg ggttcttcgg tgctggtcag
360 tgctgctcgg ctcgggtga ggacctcgag gcccaggtag cgccgtcctt cgatccattc
420 gtcgtgttgt cggcgagga cggctccgac gaggcggatg atcgaggcgc ggtcggggaa
480 gatgcccacg acgtcggttc ggcgtcgtac ctctcggttg aggcgttcct gggggttgtt
```

108

540 ggaccagatt tggcgccaga tctgcttggg gaaggcggtg aacgccagca ggtcggtgcg
600 ggcggtgtcg aggtgctcgg ccaccgcggg gagtttgtcg gtcagagcgt cgagtacccg
660 atcatattgg caacaactg a
681
<212> Type : DNA
<211> Length : 681
      SequenceName : gi_GDC_MTUB_2828257
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2828257

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgggatcgc tcaccgtgtt caccagctcg gcgaggatgt cgcgcacagc ggccaacacg
60 tcggcgcgcg cactgcacag catgaccacc gggtcgggcg ggaagagcag aatgctgaac
120 acgatagcca gcccaccacc gaccagcgcg tcgaagaggc gttcgaaaac cacactgccg
180 ttggacgcga agaccaagac cagcaccgcg gagacggcgg cctggttgat gaacattaag
240 ccttgcgcga ccaacccgcg tgcgcacagc accgcgaccg acaacgcgat gaacaccacc
300 acacccatgg cgatcggtcc ggaaccaagc agagcatgca cgccagcacc cagcacgatc
360 cccagcgcca ccccgacgat catctgttgg gcacgtcgtg cgcgcagcac gttggtcgcc
420 gacatgcaca ccacagccga aatcggcgcg aagaacgcct gcggatggtt gaacacgtca
480 tgggtgagat accacgcgag gccggcgacg accgatgtct gggtgatcgg ccacagcacg
540 gtgcgcaacc gttgggcgac cgcacggccg ccgcaggccg tcctgactag cagcgaagcg
600 ctcatgaacg cctatttatt cacactcggg tgcgacgtcg taaccgcaaa gatctggtca
660 tgcctgctgg acccgcttgg gctgggcatc tattccggac tccttacgtt gctgagcggt
720 aatgggcgcc ggcgcgtcgg tgagcggatc gacgccgccg ccggtcttcg ggaacgcgat
780 cacctcacgg atcgagtcca tcccggccag cagcgcggtg gtccggtccc acccgaacgc
840 gattccgccg tgcggcggtg cgccaaacat gaacgcctcc aacaggaatc cgaacttttc
900 ctccgcctcg gccttgtcca ggcccatcac cgcgaacacc cgttcctgga tatcacggcg
960 gtggatacgc accgagccgc caccgatctc gtggccgttg cagacgatgt cgtacgcgtc
1020 ggccagcacg ctgccggtat cggattcgat gcggtcctcc cattccggtt tcggcgcggt
1080 gaaggcatgg tgcaccgcgg tccaggcccc cgagccgacc gcgacctcac cggcggcggt
1140 cgcttcgtcg gccggctcga acagcggcgg gtcaacgacc cagacgaatg cccacgcatc
1200 ggggtcaatc aggcccagcc ggttggcgat ctcgacgcgg ccgcgcccca gcagtgcccg
1260 cgacgatttg accggaccgg ccgagaagaa gatgcaatcg ccgggtttgg ccccgacatg
1320 gtcggccagt ccggtgcgct cggcctcggt caggttttg gccaccggac cgcccagcgt
1380 gccgtcttcg cgaccagca cgtaggccag tccgcggtgg ccgcgctgct ggcccagtc
1440 ctgccagccg tccagcgtgc cccgcggctg cgacgccccg ccaggcatca ccaccgcgcc
1500 cacatacggt gcctggaaga cacgaaatgt ggtgtcggag aagaaatccg tgcattcgac
1560 gagctccagc ccgaaccgca ggtcgggttt gtccgtaccg aatcggcgca tcgcttcggc
1620 atagccgatc cgcgggatgg gcgtcggaat ccggtagcct atcagcgccc acagctcggt
1680 cagaacttcc tcggagatcg cgatgatgtc ctcggcgtcg acgaagctca tctccatatc
1740 gagctgggtg aattcgggct ggcggtcggc gcggaagtcc tcgtcgcggt agcagcgggc
1800 gatctggtag tagcgttcca tccccgccac catcagcagc tgcttgaaca gctgcgggct
1860 ctgcggtag
1869
<212> Type : DNA
<211> Length : 1869
      SequenceName : gi_GDC_MTUB_2895354
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2895354

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgatcggcg atttcgccga gatgctcggc ggccaggacg gcgtcgctga gttggtccaa
60 cacgtcgctg tgcacccgtt tgatggcgtt gatgagctcg tcgaggcgga cggggtaggc
120 ggtgggtgtg ggctccggca tgacgtcaac agtaggttga cgttatgcat tgtgtcgacc
180 gtgattggct cgtagtgggg ttctgcagcg ctgccaggcc gctgcgggca gggtggcgcc
240 gatcgcggcc accaggccgg cgtgggcgtc gctggtgacc agcgcgaccc cggacaggcc
300 gcgggcgacc aggtcgcgga agaacgccag ccagccggcc ccgtcctcgg cggaggtgac
360 ctggatgccc aggatctctc ggtagccctc ggcgttgacg ccggtggcga tcaaggtgtg
420 caccccgacg acgcggcctg cctcgcgcac cttgagcacc agggcgtcgg cggcgaggaa
480 ggtatacggg ccggcatcga gcgggcgggt ccgaaacgcc tctacggctt cgtcgagctc
540 tttggccatg atcgacactt gcgacttgga aagctttgtc acaccaagtg tttcgaccag
600 gcgctccatc cggcgagtgg atactcccag caggtagcag gtcgccacca cgctggtcag
660 tgcgcgttca gctcgcttgc ggcgctgcag cagccagtcc gggaaatagc tgccctggcg
720 cagcttgggg atcgcgacgt cgatggttgc ggcacgggtg tcgaaatcac ggtggcggta
780 gccgttgcgc tgattggacc gctcatcgct gcgttcgcgg tagcccgccc cgcacagggc
840 gtcggcttca gcccccatca aggcggcgat gaacgtcgag agcagcccgc gcagcagatc
900 cgggctcgcc tgtgcgagtt ggtcagccag aagctgctcg gtgtcgataa gatgagaaga
960 ggtcattgcg tcatttcctt cgattga
987
<212> Type : DNA
<211> Length : 987
      SequenceName : gi_GDC_MTUB_2983047
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_2983047

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttggatgagc cggcgcaccg cgctcgcccg aaagggaacg gagccaatca tgacggcgct
60 caaccgtgct gtggcatcgg cgcgtgtggg aaccgaggtg atccgcgtgc gcgggctcac
120 cttccgctac ccaaaggcgg ccgagccggc ggtgcgtggc atggagttca ccgtcggccg
180 cggcgaaatc ttcgggcttc taggtcccag cggcgcgggc aagtccacca cccagaagct
240 tctcatcggg ctgctgcgcg accacggcgg ccaggccacg gtgtgggaca agagccggc
300 cgagtgggga cccgattact acgagcgcat cggggtctcc ttcgagctgc ccaaccacta
360 ccaaaagctc accgggtatg a
381
<212> Type : DNA
<211> Length : 381
      SequenceName : gi_GDC_MTUB_3005316
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3005316

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgatccctc aaatgacggt gtcctgcccg cccccgtcga cttctgagcg cgaagagcag
60 gcgcgggcac tgtgcctgcg cctgctcacc gcgcgatccc gcacccgcgc cgagttagcc
120 ggccagctgg ccaagcgcgg ctaccccgaa gacatcggca accgggtatt ggatcggctg
180 gccgccgttg gcctggtgga tgacaccgac ttcgccgaac aatgggttca gtccaggcgg
240 gcgaacgcag caaagagcaa gcgcgcgttg ctgccgagc tgcacgccaa gggcgtcgac
300 gacgacgtga tcaccacggt gctcgggggc atcgacgccg gtgccgaacg ggggcgggcg
360 gaaaagctgg tacgggccag gctgcggcgg gaggtgctga tcgacgacgg caccgacgaa

420 gcgcgggtga gccgcaggct ggtggcgatg ttggcgcgcc gtgggtacgg ccagaccttg
480 gcgtgcgagg tggttatcgc cgagctggcc gccgagcggg agcgccgacg cgtctaa
537
<212> Type : DNA
<211> Length : 537
    SequenceName :ˑgi_GDC_MTUB_3048559
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3048559

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttggtgacga ctctggcgcc gatcttggac agtgcatcga tgactccgaa gaccgcctcc
60 tcgttgccgg ggatcagcga cgacgacaac acgatgagat caccagcagt caacgtgatg
120 ctgcgatgct ccccacgcga cattcgcgac aacgccgaca tcggctcgcc ttgggtgccg
180 gtggtgatca acacaacttg gtcgggcgcc atcgtttcgg cggcggcgat gtcgatgaga
240 tcggaatcag ccactcgtag gaagcccagt tgccttgcga cgcgcatgtt gcgcaccatc
300 gatcggccga cgaacgacac tcgccggccc aatgccactg cggcatcgat gatctgctgt
360 acccgatcca cgttggaggc gaaacacgca actatcaccc gtccgtcggc accccggatg
420 agccggtgca gcgttgggcc cacttcgctt tccgatggcc cgacaccggg gatctcgggcg
480 ttcgtcgagt cgcacagcaa caggtccacg ccggtgtcgc cgagccgcga catgcccggt
540 agatcggtgg gacggccgtc cggtggcaat tggtcgaact tgatgtcgcc ggtgtgcagg
600 atggttcccg cgccggtata caccgcgatg gccaacgcgt ccggagtgga atggttgacg
660 gcgaagtact cgcactcaaa cacgccgtgc cgggtgctct ggccctcgcg gacctcgacg
720 aacaccggtg ttatgcggta ctcacgacat ttctctgcaa ccagagccaa ggtgaacttc
780 gagccgacga ccgggatgtc gggtcgcagc ttgagcagaa acggaatcgc cccgatgtgg
840 tcctcgtgcc cgtgggtcaa caccagcgcc tcgatgtcgt caagccggtc ttcgacatgg
900 cgcatgtccg gcaggatcag atcgacaccg ggctcgtcgt ggccaggaaa caacacaccg
960 cagtcgataa tcaacagtcg gcccaggtgt tcgaaaccg tcatgttgcg ccgatttcg
1020 ttgatgccgc ccagcgcggt gacccgcaac ccgccggagg tcaggggacc tggcggggga
1080 aggtctacat ccacttctgg gccacccttt ggctcacctt tagatcaccg aagcaccgag
1140 gccgcgcgca tgtcggcggc caacgcgtcg atctgctccg gtgtcgcggc cacctggggc
1200 agccggggat caccgacgtc gatgccctgc agccgcaagc ccgccttgga caacgtcacc
1260 ccacccaggc ggctcatcgc gttgcacagc ggggcgaccg caatgttgat cttgcgggcg
1320 gtggcgatat ccccagaacc gaaggcggac aacaactctc gaagctgccc ggctgccagg
1380 tgggcaatca cgctgatgaa gcccgtggcg cccatggcca gcagggcag gttgagcgcg
1440 tcgtcgccgg aatag
1455
<212> Type : DNA
<211> Length : 1455
    SequenceName : gi_GDC_MTUB_3065095
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3065095

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtccaaga gatcggatgg gccgagcact ggcaatgcga ttcgtgctcg gcatcgcatc
60 agcgtgatga ctgcgcagcg atcaacctcg cacgctacga ggacaccagt agcgtcgtcg
120 gcccagttgg ggccgccgtc aagcgtggag ccgaccgtaa gacccggcct ggccgggctg
180 gtggccgtga agcgcggaag ggaagcagcc gcaaggctgc cgaacaaccc cgagacgggg
240 tgcaagtcgc gtgaccacta a
261

```
<212> Type : DNA
<211> Length : 261
      SequenceName : gi_GDC_MTUB_3100192
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3100192

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggcaacga agaacgcggc atggccttca tctacaagct gctcgaacta ctcgccgaac
60 gcgacgatcg aatcacaaag gccagatggg tgtacttcct cacgcgcatg cgtaacccca
120 ccggtgacac agcgcctttt cagcagtttg ctaaccggct acaccaatgg ttccaagatc
180 cgacagacgc caagcaactc aagaccgcgc tgcacctcta catctatcgc actcgcaagg
240 aggagtccga atgagcgtca tccaagacga ctatgtgaaa caggccgaag taattcgcgg
300 cctgccaaag aaaaagaacg gcttcgagct gaccacaacc cagctgcggg tgctactcag
360 cctgaccgca cagctcttcg acgaggcgca gcagagcgcc aaccccacgc tcccgcgtca
420 gctgaaggag aaggtccagt acctgcgggt ccggttcgtc taccagtccg ggcgtga
477
<212> Type : DNA
<211> Length : 477
      SequenceName : gi_GDC_MTUB_3129118
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3129118

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtcggcgc ctgacgtgcg gctgaccgcc tgggtgcacg ggtgggtgca gggagtcggt
60 ttccgctggt ggacccgctg ccgagcgttg gagctcggcc tgaccggtta cgcggccaac
120 cacgccgacg gacgcgtgct ggtggtcgcc cagggtccgc gcgctgcgtg ccagaagctg
180 ctgcagctgc tgcagggcga cacgacaccg ggccgcgtcg ccaaagtcgt cgccgactgg
240 tcgcagtcga cggagcagat caccgggttc agcgagcggt aa
282
<212> Type : DNA
<211> Length : 282
      SequenceName : gi_GDC_MTUB_3237815
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3237815

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgttgcacg acgtcgtcca cggcagacga tgtagtgaga atggccaccg gcgacgaatc
60 actcagtacc gaatcggaac gttcatcggt aacgccgcct tgtggaaccg aaagcggcac
120 ggcgatgcgc ccggcctgca acgcgccgag aaaggcgacg acgtactcga gtccctgcgg
180 agcagagatc accacgcggt cacccgtgga accacaacgg ctcagctcct gtgccacatt
240 cagcgttcgc cgatacagct gcgaccacgt cagggttatc gcaacgccgt cccagtcctg
300 ttcgtaatcc ataaacgtga aggccgggtc atggggttgc agacgcgcac acgcgcgcaa
360 cgcagcggga agggaacgca cactcatggg catcacgtta ccggccacgc ttggagttgt
```

```
420 cgcagtcgcc gtcggggtgt gctcgcgctc cgcggtctta gccaagtcgc atctggccag
480 ctcagcaggg gtttgccggc tcgccatggg tccaccatcg gacacggtcg gatgtga
537
<212> Type : DNA
<211> Length : 537
      SequenceName : gi_GDC_MTUB_3283182
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3283182

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis'
<400> PreSequenceString :
atgcccacca ccaaagccac ccagcgccgt gatgtttcca ccgagatcgc ttacctgaca
60 agagcattga aagctcccac cctgcgtgag tcagtgtccc ggctggccga tcgcgcccgc
120 gccgagaact ggagccacga agaatacctg gccgcctgcc tgcagcggga agtgtcagcc
180 cgggagtccc atggtggtga gggccgcatc cgcgccgccc gcttcccggc tcggaagtcg
240 ttggaagagt tcgactttga gcatgctcgt ggcctcaaac gcgacaccat cgcacatctg
300 ggcaccctgg atttcatcac cgcccgcgat aacgtcgtgt ttttgggccc cgcctggcac
360 cgggaagact catcttgcgg tcggcctggc gatacgcgcg tgtcaggccg gtcatcgggt
420 gctgttcgcc accgccgccg aatgggtagc acggctcgcc gaggctcacc acgccgggcg
480 catctacgcc gaactcaccc ggctttgccg ctatccgctc ctggtggttg a
531
<212> Type : DNA
<211> Length : 531
      SequenceName : gi_GDC_MTUB_3289702
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3289702

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcagtggg ggtaccgccc gcttgcgggg gacgaagcga tgaggtgggg gtaccgcccg
60 cttgcgaggg agagcggcgc acttgacccg gatcatcggc ggtgtcgccg gaggccggcg
120 cattgccgtc ccaccacgcg gaaccagacc taccaccgat cgggtgcgcg agtcgctatt
180 caacatcgtg actgcgcggc gggatctgac cggtctggcg gtgttggacc tctatgcggg
240 ttccggcgcg ctgggctgg aggcgttgtc gcggggagcg gcgtccgtgc tgttcgtgga
300 gtccgaccag cgcagcgcgg ccgtcattgc gcgcaacatc gaggccctag gtctctccgg
360 tgcgacgctg cgccggggcg cggtggcggc cgtcgtggcg gccgggacca cgtccccggt
420 ggatctggtg ttggccgacc cgccctacaa cgtcgactcc gccgacgttg a
471
<212> Type : DNA
<211> Length : 471
      SequenceName : gi_GDC_MTUB_3319076
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3319076

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
```

```
      ttgggtgggg ttgccagcac tcggcaggca tccgttcgcc gttggtctgc cgttcacccc
  60  ctggatgcct cgccggcgtt gccccgtccc ggtcaacgat gtgcgaccgc tcgcgcggtc
 120  gcgggcccta ccccgagctg gcgtgcggcc gtcaggtcgg cggggggtgtc gacatcgcag
 180  cgcaggcccg gccaggctcc tgtcagctcg acagcgcccg aacggcggtg ccgcgcggac
 240  gaatccggcc cgaaccgcgg gtgcagcgcg gtgccgaacg cacacagtac cgcggtgccg
 300  gtcccaagcc ggtcggcgac gaagctgcgc cgatggtggc gtgcggccga gattgcctcg
 360  gcgagttcct gtgtctgtaa tgccggcaaa tcgccttgca gcacaacgat gttggaggcc
 420  ccttcggcaa ccacgcgttc ggcagcggtg atggcggtgt tcagtgggtc gggatcgtct
 480  tcgggtgtcg ggtcggccag tacatcggcg cccagcccgg ccgccgcagc cgccgcggct
 540  tcgtcggggg tgataacagt gatcgagcgc agtgaaccga cacccgccgc ggcggtcaac
 600  gtgtcgacga gcatggccag caccacgttc tcgcgagtct gcgccgagaa caccggggcc
 660  agcctggttt tggccgcggc caagcgcttg acggcgatga tcaagccgat atcgccgtcg
 720  tccggtgtgc cgctcatgaa gtcatcctgc cagcgtcgat ccacgcggca cacttcgacg
 780  gcattgccgc cacggtcgtg gccggggccc aggcacggtc ccgacggcaa ccgcggcgca
 840  gattag
 846
<212> Type : DNA
<211> Length : 846
      SequenceName : gi_GDC_MTUB_3339006
      SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3339006

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
      ttgcgcggca ggttgatccg atacgcggtg ttgttgtctc cgagcttgcc gctacgtccc
  60  agcgcgtcgg ccaccggctt ccagtcggca tcggtggtgg tcaccgccga acgagctttg
 120  ccggcgtggc cgctgcccgc tccacccttg gagcccgaac tgcacgccgc cagtatcacc
 180  gccgccgcgg tggtgatcgc gacgattctc ccagcatgtt tggcgcccgc catgcgcgtt
 240  ccctccatcc gttgcatcca cggcgtggat ggcagttcgg ttagccatgg tctatcgggt
 300  gattatgaaa ccacgatgaa gctcgatcgc accgatccgg gcacggccag acgtcctcat
 360  cgacgccctg ggcgcgtatc tgctggccgc cgcggctctt cgacccgtgg aacgcatgcg
 420  catccgcgcc gcgggcatca gcgccaccga cccacatgcc cgtctgccat tgccactggc
 480  tcgagacgaa atccggtatc ttggaacaac attcaacgac cttctgcagc ggctgcaaga
 540  cgcgctcgag cgagaacgtc aattcgtcag cgatgcgggc acgaacttc gcaccccctt
 600  agcctcctga ccaccgaact cgaactcgcc ctgcggcgtc cacgaagcaa ccccgaactg
 660  ctcgccgcaa tccgctcggc tctcgcggaa accaccgaca ccgcgcgcac caccggcggc
 720  accgggcttg gactggccat cgtcgacacc ctcagccaac gcaaccacgc cagcgtcacc
 780  gcccgaaacc gcgccgcagg cggtgccgaa atctccctcc ggcttgctct ggctga
 837
<212> Type : DNA
<211> Length : 837
      SequenceName : gi_GDC_MTUB_3356995
      SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3356995

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
      ttgcttgggc tgcccgaccc ccgccccgtc ccacgcaacc cggctgcccg tcgtcgggcg
  60  acatcccggt ctctatcggc ggacccgagc agccgcccgg ctagccagtc gcggccaagg
 120  ccagggacgt ggtgtacgag tgaaggttcc tcgcgtgatc cttcgggtgg cagtctaggt
 180  ggtcagtgct ggggtgttgg tggtttgctg cttggcgggt tcttcggtgc tggtcagtgc
```

```
240 tgctcgggct cggggtgagga cctcgaggcc caggtagcgc cgtccttcga tccattcgtc
300 gtgttgttcg gcgaggacgg ctccgacgag gcggatgatc gaggcgcggt cggggaagat
360 gcccacgacg tcggttcggc gtcgtacctc tcggttgagg cgttcctggg ggttgttgga
420 ccagatttgg cgccagatct gcttggggaa ggcggtgaac gccagcaggt cggtgcgggc
480 ggtgtcgagg tgctcggcca ccgcggggag tttgtcggtc agagcgtcga gtacccgatc
540 atattgggca acaactga
558
<212> Type : DNA
<211> Length : 558
      SequenceName : gi_GDC_MTUB_3381198
      SequenceDescription :

Custom Codon
-------------
Sequence Name : gi_GDC_MTUB_3381198


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgattttct gggcaaccag gtactgcacg atctggttgc cgccttcacc ctcgtcggtg
60 accttctccc cggcagtctt ggccggtttg ggcgtcgacg ccagcacggt ggatccggcg
120 ttggccagcc ccacctcgtc gctctcgaca ccgatctcgg ccagggtcag cacggtaact
180 tccttcttct tggcggccat gatgcctttg aaggacggga agcgcggctc gttgatcttc
240 tcgttcacgc tgatcaccgc gggcagcgtg gcctcgaggg tgaatacgcc ctcatcggtc
300 tcacgctcgc cggtgatctt gccgccctcg atcgacactt tgcgcaggtg ggtgagctgc
360 ggcaggccca ggtactcggc gatgatggcc ggcaccgcac cgcccacccc gtcggtcgat
420 tcgttgcctg cgatcaccag ctcggtgccc tcgatggtgc ccaacgcgcg cgccaaagcc
480 cacccggttt ggatgacgtc cgagccgtgc atgccgtcgt cctttaggtg gacggccttg
540 tcggcaccca tcgacagcgc cttgcggatc gcctcggtgg cgcgctcggg gcccgccgtc
600 agcacggtta ccgacccttc gatgccgtcg gcggcctctt ctcccgaat ctgtagcgct
660 tcctccacgg cgcgctcgtt gatctcgtcc agcaccgcgt cggcggcctc gcggtccagc
720 gtgaaatcgc cgtcggtcag cttgcgctcc gaccaggtat ctgggacctg cttgatcagg
780 accacgatgt cgtcatgac tgtggttcgt cctcctcgaa ggcggcccgc agcgctcgac
840 tgcggaacct cggtcacacg ttttgcaacc gcacagcgat attactattc ggtaagttcg
900 cgtggtgcgc cctcacacca tagcgggtgg tag
933
<212> Type : DNA
<211> Length : 933
      SequenceName : gi_GDC_MTUB_3388071
      SequenceDescription :

Custom Codon
-------------
Sequence Name : gi_GDC_MTUB_3388071


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgctctcct cctggccaag gccagggacg tggtgtacga gtgaaggttc ctcgcgtgat
60 ccttcgggtg gcagtctagg tggtcagtgc tggggtgttg gtggtttgct gcttggcggg
120 ttcttcggtg ctggtcagtg ctgctcgggc tcgggtgagg acctcgaggc ccaggtagcg
180 ccgtccttcg atccattcgt cgtgttgttc ggcgaggacg gctccgacga ggcggatgat
240 cgaggcgcgg tcggggaaga tgcccacgac gtcggttcgg cgtcgtacct ctcggttgag
300 gcgttcctgg gggttgttgg accagatttg gcgccagatc tgcttgggga aggcggtgaa
360 cgccagcagg tcggtgcggg cggtgtcgag gtgctcggcc accgcgggga gtttgtcggt
420 cagagcgtcg agtacccgat catattgggc aacaactga
459
<212> Type : DNA
<211> Length : 459
```

SequenceName : gi_GDC_MTUB_3482312
          SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3482312

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgatcagat cgatcgatcg ctgggggtcc gctgccgggg gggcggtcgg cacgcccggt
60 gggaccgact gtaatggccg ctcctcccac ccagctcggt ctgcggcgac gaacacatcg
120 atctcggccc agggcgccgc gggtccctgg gtcaagaatc gggggcgttc cagttttccg
180 gtggcctcat gcagccgcac cgccgccgag acgacctcat catgcctagg ctccggcgcg
240 ccggcgacga acgtgtctgc ccgccaacca gacaccacgt accggccgtc ggtcgatcgg
300 acgggccgag ccaggcgtac gccgtcgacg aacaacgtct cgcgcacccg gccgaccag
360 gccgcgcggg cgttgtcggc caccatcgac aacaccacct cgccgcatcg ccagccacct
420 tcccaaccgg cacccaacag gatgggttgc gcacctgcca aaccgaacgc caccaacacg
480 tgctcgggcg cgggctcgac attcacaccg gtcagcctag tagagcccat cggggtgtat
540 tgggcctgta tcggtcctag tacatcacca tgtcgggctg catctgcttg gcccacgcga
600 cgatcccacc ctgcaggtgt accgcgtcgg agaaaccggc tttcttga
648
<212> Type : DNA
<211> Length : 648
          SequenceName : gi_GDC_MTUB_3581973
          SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3581973

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgatgttct gtgcgtcgcg gaaagagatg gcgatgtcga attcgtcttc tagctcggtg
60 atcaactgga acagcttgag cgagtcaaaa cccaggtcgt cgacgagtac ctggttcgcg
120 gtgatgccgc ggtcggttcg caagatccgt tggatggtgg cgttgatggc ctctttcata
180 gcgcggctcc ttgcggggtc aggtcctcgg caaggccggc aaacacgtgc aaggcccggt
240 cgaggtcaga ttgtcggtgg tcggctaggt agctggtgcg gaatcccgaa cgctcctccg
300 gcacggctgg gggggccacc gggttcacat acaccccgga gcgcatcagc cgcagatagc
360 ccgcatgcgc cacggtcggg ttgcccagga tcaccggcac gatcgcggtt ccgtgatact
420 cggcctgata gccctgccgt gccaggccgg tggccatgta ctcggccgcg ccagcaccc
480 gagcccgccg gtcgggttca cgccgactga
510
<212> Type : DNA
<211> Length : 510
          SequenceName : gi_GDC_MTUB_36276
          SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_36276

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcggtgta gggcggcgtt gagctggcgg ttgcccgagc ggctgagccg catctggccg
60 gcggtgttgc ccgaccacac cgggatggga gccactgcgg catggcaggc gaaggcggct

116

```
120 tcgcttttga accgggtcac tccggcggct tcgccgacga ttttggctgc agtcagctcc
180 gcgcagccag ggatttccag cagtgcgggg gcgacctggt ggactcgggc gctgatgcgc
240 tgggctaggg tgttgatctc gccggtgagc cggatgatgt cggtcagctc ggcgcgcgcg
300 agttcggcga ccaatcctgg ctgggtgtcc agccaggtcc gcagggcctg ctggtgcttg
360 gcggcatcgg cgagcgtgc tgccggtggc cgctcgggat cgagttcatg gacgagccag
420 cgcaaccggt tgatcgccga cgtgcgttgg gccacaagga catctcgacg gtcagtcaac
480 aacttcaact cccgcgacgt ctcgtcgtgg gtggccaggg gtaggtcggt ttcacgcagc
540 accgcccgcg ccaccgccag cgcatcgatc ggatccgact tgccccgact gcgcgccgac
600 ttgcgggtct gggccatcag cttggtgggt acccgcacca cctgctggcc ggccgccagt
660 aggtcacgct ccagacgcgc cgacatgttg cggcagtcct cgatgcccca gatcagctcg
720 aggccgaact gttcacgggc ccacatgatg gctgtggcgt gcccggccgt ggtggccttg
780 acggtcttct caccgagttg gcgacccact tcgtcggtgg ccacaaaggt gtggctgtac
840 ttgtgcgcat cggttccaac aacaaccatg gtggttgcct ctgaaccgcc ccggtga
897
<212> Type : DNA
<211> Length : 897
      SequenceName : gi_GDC_MTUB_3711717
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3711717

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgccggatc tcctcgagtt tgcggccctt ggtctccggc gcaaagcggt acacgaccac
60 gaacgcgacg acggcgaacg tgccgaagac cgcgaaaacg cctgcgccgc cgagcacacg
120 cagcatggtg agcgagaagg cggcaacgat cgcgttggcc gtcagtgtcg aggtgagcat
180 cgggctcgat cccatcgacc gcagccggga cgggaagctc tccgcggcgt acacccagac
240 cagcgagccg aatccgaagt tgaacccgat gatgaacagc agcacgccgg cgaaccccaa
300 caccagcccc gtgccaccat cggagtcgtt ggcgaatacg gtgatcagca cggcatctgc
360 ggtgatcatc gtcgcgatgc cggacaacag gatcgggcga cggcccagcc gatcgaccag
420 aaacagcgag gcacacaccg ccgccaagcc ggcgacttgc accatcgcgg gcagggcaag
480 catcgcgaaa tagcccgcga gcccatggc ggcgaaaagt cgcggactgt agtagatgat
540 cgcgttgatc ccggtgatct ggacgaggaa gccgagcgcg atgacgaaca gcgtggcccg
600 cagatacggc cgccgcacca tttcgccgat accgccgccg cgttcgtcga ccgcggccgc
660 catatcggcc agctcggcat cgatgtcggc ctccggctgg atccgccgca gcgcgctacg
720 cgcgtcggcg atccggccct tgagcagata ccagcgggcg gtatcgggca tgcgccacaa
780 caacggcaac agcagcgtgg ccggcgcggc ggccagcccg aacatcgcgc gccagccgtg
840 cgatccggcc aacaggtagc cgaccaggta accgacgacg atgccgctaa gcgtcgccag
900 ctgatacgcg gtcaccacg acccacgcac cgccgccggc gccgactcgg ccacatacac
960 cggcaccacc accaccgaca ggccgattgt cacacccagc agcagacgcg ccaccaccag
1020 catcggtacg gacaccgagg tcgcgccgag cagggcgaac actgcgtagc cggcgacgat
1080 gagcaccacc gatttcttgc gtccgatcgc gttggcgagg atgccgccgc caagcgcccc
1140 ggcgatctgg ccgagcaccg ccgtggtggt cagcaactcc tgttctcgag tggtgagttc
1200 gaattcctcg ctgagagaca gcaacgcacc cgcgatggcg gaaaggtcgt acccgtagag
1260 gacgccgacg ctggcggcgg tgagcccgac gaggagcgcc cggcgccccg atctggtcag
1320 ttgaccggta ccggggcgct cagcacgtcc accacgcggt cggggtcgtc gggcgccccg
1380 gcgggcgtga accccgcgtc ctggtatagg gctgtagtca tttcgatgag gctgccacag
1440 cgtcgtcacg cggtcaaccg ctggtcaagc cccgatttcg gtgccgacca aggccggcta
1500 ggatgcccgc cgcaaacga cgccgaaggt atcggggtaa gcagctag
1548
<212> Type : DNA
<211> Length : 1548
      SequenceName : gi_GDC_MTUB_3716987
      SequenceDescription :

Custom Codon
------------
```

Sequence Name : gi_GDC_MTUB_3716987

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgtctgacg ctacgacagt gttgttcggg ctgccaggag cacgggttga gcgtgtcgag
60 cgccgcagtg acgggacccg ggtggtcgat gtgatcaccg atgagccgac ggcggcggcg
120 tgcccgtcgt gcggggggtgg tctcgatatc agtgaaggaa tacgcggtta cctcaccgaa
180 agatctacct tatggcgaag accgcatcat ggtgcgctgg aacaaaattc gctggcgatg
240 ccgagaagac tactgcaagc tggggccgtt caccgaggcc atcacccagg tacctgcccg
300 cgtccgcagc acgctgcggc tgcgtcggca gatggccaag gcgatcgggg atgcggcccg
360 ctcggtgggc cgaggtcgcc caggctgacg ccgtgtcgtg gccgacggca catcgggcgt
420 ttgttgccta cgccgagacg ggtattgacc gagccgttgc ccaccccggt gctgggcgtt
480 gaccagacac ggcgaggaaa acccagatgg gagcgctgcg ccaagactgg ccggtgggta
540 cgggtcgacc cgtgggatac cgggttcgtc gacctggccg gtgatcaggg gtttatgggg
600 cagcatgaag gccgcggcgg cgcggcggtg ctggcatggc tgcaagcgcg cacaccgcag
660 ttccgggaga gcatccagta cggtggccat cgaccccgcc gctgcctacg cctcggcgat
720 ccgcacgccc gggctgctgc ccaacgccaa gctcgtcgtc gaccacttcc atgtgaccac
780 gctggccaac gacgcgctga ccgcggtgcg ccgccgggtg acctgggcgt tccacgaccg
840 gcgcggccgc aagatcgacc cgcagtgggc caaccgacgt cgcttgctga ccgcccggga
900 acgcttgtcg gacaaaagct tcgccaaaat gcggaatcgg atcaacgccg tcgacccccg
960 cgcgcagatt ctctcggcct ggatcgccaa agaggagctg cgcaccctgc tgtcgaccgt
1020 gcgcaccggc ggggacccccc acctggcgcg ccatcaccta caccgcttcc tgcctggcgc
1080 atcgactcgc agatccccga actgctcacc ctggccacca ccattgacta g
1131
<212> Type : DNA
<211> Length : 1131
     SequenceName : gi_GDC_MTUB_3754581
     SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3754581

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcaggcat tgcccgaaag ccagctgcca gagctggccg tgcagatgcg tcggcggctc
60 atagaaacag tgacggctac cggtggccat ctcggcgcgg gacttggcat ggtagagctg
120 accatcgcat tgcatcgggt gttcacctcg ccacacgaca tcggtgttcg acaccgggca
180 ccaaacctat ccgcacaagc tgctcaccgg ccgcggtaa
219
<212> Type : DNA
<211> Length : 219
     SequenceName : gi_GDC_MTUB_3794808
     SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3794808

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgtcttcag aggggggttg gcccaacgtc ggaaacctcg cgcgcagcgc atcaatgaca
60 tcggcagttt catcaagtgc cagg'gttgtc tgggtcagat acgatagctg ggtaccctcg
120 ggcaggttca acgctgccac atcagcgggt gtctgcacca ataatgttga ccgcggagcg
180 acgccaagcg tgccttcggt ctcctcatgt ccggcgtgcc cgatgaagac caccgtgtca

118

```
240 ccgcgcgcgg caaaccgtgc ggcttcagcg tggactttcg ccaccagtgg gcaggtcgcg
300 tcgacgacct gcagtccccg ctcatcagcg cccgcgcgca ccgccgggga aaccccatgc
360 gcggagaaca ccacgaccgc ccccggcggc ggcggatcgg gaatctcgtc gagatcctcg
420 acgaacactg ctccccggtc ccgcaactcg gcaaccacaa cagtgttgtg cacgatttgc
480 ttgcgcacat acaccgggcc ttcggccacg tcaagcactc gcttgaccgt ctcgatagca
540 cgctctacac cggcgcaaaa cgaccgcggc gacgccaaca gcaccgtgac ttcaccgaa
600 gcgtatccct gtgcgaccgg tcccacgaac acctcagcca tcagcactcc cggcgacata
660 tcagttgcga caacgcgatc aggtctgggg atcgcaccgc atcgggcagt gccgcaatag
720
```
<212> Type : DNA
<211> Length : 720
　　　SequenceName : gi_GDC_MTUB_3796793
　　　SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3796793

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
```
    ttgcctgggc atcgtcgggg cacgtcggct tcaagggttc ccggaaatcg accccgtttg
 60 cggcccagct ggccgcggag aacgccgctc gcaaggccca agaccacggg gtgcgcaagg
120 tcgacgtgtt cgtcaagggc ccgggctcgg ccgcgagac cgcgatccgg tcgctgcagg
180 ccgccggcct ggaggtgggc gcgatctcgg atgtcacccc ccagccgcat aacggtgtcc
240 ggccccccaa gcgccggcgc gtctaggaga gaagatggct cgttacaccg gacccgtcac
300 ccgcaaatca cggcggttgc gcaccgacct cgtcggtggc gaccaggcct tcgagaagcg
360 tccctacccg cccggccaac acggtcgcgc gcggatcaag gaaagcgaat atctgcttca
420 gctgcaggag aagcagaagg cccgtttcac atacggcgta atggaaaagc agttccgccg
480 ctactacgaa gaggccgtgc ggcagcccgg caagacgggt ga
522
```
<212> Type : DNA
<211> Length : 522
　　　SequenceName : gi_GDC_MTUB_3879013
　　　SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3879013

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
```
    gtgggacgcc gtgatcgcgg tgcacctgcg cggccatttt ctgctcaccc gcaacgccgc
 60 tgcctactgg cgggacaaag ccaaggatgc cgaaggggga tcggtcttcg gccggctcgt
120 caacacctcg tcggaggcgg gtctggtggg cccggtgggg caggcgaatt acgccgccgc
180 caaggctggc atcaccgcgc taaccctgtc ggcggcgcgg gcgctcgggc gctacggcgt
240 ttgcgccaat gtgatttgtc cgcgggcgcg caccgcgatg acggccgatg tcttcggcgc
300 cgcacccgat gtcgaagcgg gccagatcga cccgctgtcg ccgcagcatg tggtaagcct
360 ggtccagttt ctggcgtccc cggctgccgc ggaagtcaac ggtcaggtgt tcatcgtcta
420 cggtccgcag gtgacgctgg tgtcaccgcc gcacatggag cgccggttca gcgcggacgg
480 cacgtcctgg gatcccaccg agctcaccgc gacgctgcgg gactactttg ctggtcggga
540 tccggaacag agcttttcgg cgaccgatct gatgcgtcag tgacccgtgg atataggcgg
600 ccgattattg gaatcggtgt ccgaatcacc acgccaacat ag
642
```
<212> Type : DNA
<211> Length : 642
　　　SequenceName : gi_GDC_MTUB_3921024
　　　SequenceDescription :

119

Custom Codon

Sequence Name : gi_GDC_MTUB_3921024

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgccttgga cggcatgttg ctcccttat tcgaacgaca accggaccaa acccagcccg
60 gtgaagtcgg cgacaaactc gtcgccggcc cgcgcctcga ccgcgaacgt gcatgacccg
120 ggtaacacga tgtcgccttt gcgcagccgc acgccgaaac tctcgacctt gccggccagc
180 caagccaccg cggtcgccgg gttacccaac accgcatcac tgcggccctc ggccaccacc
240 tcgccgttgc gggtcagctt cgcatcgatc gccctgacgt caagatcggc cggcggcacc
300 cgggccgcgc ccaacacgaa gcccgccgcc gaggcgttgt cggcgatggt gtcgcagatc
360 ttgatctgcc aatccttgat cctggtgtcg atcagctcga tggcgggcac cagggcctcg
420 gtggccgcca gcacgtcgtc ctcggtgcag cccgcacccg gtaggtcggc ggccaggatg
480 aagcccacct ccacctcaac ccgcggagac aggtaccggg acgcctggac cggcgtgtct
540 tcgaacacct gcatgtcgtc gagcaggtgt ccgtag
576
<212> Type : DNA
<211> Length : 576
      SequenceName : gi_GDC_MTUB_3974481
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3974481

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggttcact ctcggcgctc atgggcgcca tcccgccgcc cgcatcgcgg catcgacgcg
60 gccaacgaac gtgccccggc ggtaccagag cagctcactg gtgaccctga tgatcgtcca
120 gcccagatcc agcaacgcgg tggaccgctc gatgtcccga gcccgctgcg ccgggtctgt
180 ccaatgctgt ggcccgtcat actcgacacc gactcgcaat tgctcgtagc ccaggtcgat
240 gcgggcgacg aagtccccgt agtcgtcaaa cactctgatc tgtgtttgcg gcttcggcag
300 accggcatcg atcaacacca atcgggtcca cgtctcctgt ggggattccg cacccccgtc
360 gatcagcggc agcaccgcac ggaggcggac caggccgcgc gcaccggtat gttcggcaat
420 gacggcctgc acgtcggcga ccttgacatc ggtcgaattc gccaacgcgt ccagccgttg
480 aacggcctgc agccgcgagg gtgtgcgccg cccgatatcg aaggcggtgc gcgccggggt
540 ggttaccgcg acaccgtcaa ccgcaaccgt ctcgtgcggc gccaatcgat ccgtgtgcac
600 gacgatgcgc ggcggaggct ttcgattggc gtgcactaa
639
<212> Type : DNA
<211> Length : 639
      SequenceName : gi_GDC_MTUB_3994808
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_3994808

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgtcgcgct accccaacag ctggcgcagg ttgaacaacc ccgatatggc ggtgcccatg
60 ttaaacaggc ccgtgttcaa gccgctccgg acggagccaa agagggtgcc cgggacgccg
120 atgttgccaa tgcccgaggt ctggccgttg atgacagtgc ccccgctggc cgtgttgaag

```
180 aacccggaga cgtcgacggc taaggggccg gtggggggtgt tgaagaagcc cgagacgtcg
240 gtgccggtgt tgccgaagcc cgagttggtc aggccgctgt cggtaatgat cccgaaaccg
300 gtgttcacat tgcccgcatt ccacgagccg gtgttgatgt tgcccgagtt cccattgccg
360 gtgttgacgt tgccggagtt gtcaaacccc gtgttgacga agcccgcgtt ccgaagccg
420 gtgtttaatt caccccgcgtt ccccaagccg gtgttgagga tgctcgcgtt cccgaagccg
480 gtgttgagaa cgcccgcgtt cccgaagccg atgttggcgt tgccggaatt cccgacgccc
540 aggttgttga ggtcgccagg caccagggta ttggctccgg tgttgaagac gccgatgttg
600 ccgctgccgg agttgaacaa gccgatgttg ttggtgccgg agttgccgat gccgatattg
660 ccgctgccgg agttcagcag cccggccagg ttgatgccca tctga
705
<212> Type : DNA
<211> Length : 705
      SequenceName : gi_GDC_MTUB_3998938
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_3998938

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgagctcaa atcatgcgat tctgcgtctg ctcgcgccct tgcggctaga tccccagaac
60 ctgggcgctg gcccacagcg cgagcaccgc catcgccagg gccgcaggca cggtgcacag
120 tcccagtcgg gtgtactcgc cgacgctggc gtcgacgttg tgccggcgca gcacgccccg
180 ccacagcagg ttagacagcg aaccggcata ggtcaggttg ggtccgatgt tgaccccgag
240 tag
243
<212> Type : DNA
<211> Length : 243
      SequenceName : gi_GDC_MTUB_4021183
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_4021183

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgtgccagg gtgtacccgc ccgattgccg ccggcaaccg acactgttgg tgtagtgacc
60 aaatcagcag tgccccgggt gggtcttgac gtgcaaatcg actacagtct tggtgaccgt
120 ccggtacccg ggcatgggac tggaacgaac caagaaacct gtgaggccgt ctgctatgga
180 gcggttcgac ggtttgcgtc cggccaggct caaggtgggg atcatctcgg ctggccgggt
240 cggcaccgcg ctaggggtcg cgctgcagcg cgccgaccat gttgtggtgg cgtgcagcgc
300 catctctcat gcgtcccggc ggcgcgcgca gcgccggctg cctga
345
<212> Type : DNA
<211> Length : 345
      SequenceName : gi_GDC_MTUB_4045946
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_4045946

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
```

<400> PreSequenceString :
atgcggcccg caaaacgggc cgaggaggag ccaggcaatc accccagagc cgggtgcagc
60 gggtcgccac catcagcccc gtggcgatcg caaaccccgc gcctggcgac aatgcggccc
120 gcaaaacggg ccgaggagga gccaggcaat caccccagag ccgggtgcag cgggtcgcca
180 ccatcagccc cgtggcgatc gcaaaccccg cgcctggcga caatgcggcc cgcaaaacgg
240 gccgaggagg agccaggcaa tcaccccaga gccgggtgca gcgggtcgcc accatcagcc
300 ccgtggcgat cgcaaacccc gcgcctggcg acaatgcggc ccgcaaaacg gccgaggagg
360 gagccaggca atcaccccag agccgggtgc agcgggtcgc caccatcagc cccgtggcga
420 tcgcaaaccc cgcgcctggc gacaatgcgg cccgcaaaac gggccgagga ggagccaggc
480 aatcacccca gagccgggtg cagcgggtcg ccactggcta gaccaacgac cggtagttcc
540 cgacggcgtc ggaaaatccg acagctgagc gttcgggtca aacacgcggt gcaccggacc
600 tga
603
<212> Type : DNA
<211> Length : 603
    SequenceName : gi_GDC_MTUB_4053033
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4053033

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcgcacta cgatcgacct cgatgacgac atactgcggg cgttgaaacg acgccagcgc
60 gaggagcgca aaacgttagg gcagctcgcc tccgaattgc ttgcgcaagc tctggcggcc
120 gagcctcctc caaacgttga catccgctgg tcgactgccg acttgcggcc ccgtgtggat
180 cttgacgaca aggacgctgt ttgggcgatt ttggaccgtg ggtga
225
<212> Type : DNA
<211> Length : 225
    SequenceName : gi_GDC_MTUB_4140236
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4140236

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgtcacgtt gtcggattca ctgtcgccgg ctagcgcttt ccgtcagaa gacgagaagc
60 ctccccgatc tccaactagc atcgagatcg ggcttgcgaa ggttgggttg caaaatggat
120 gtcatcagat gggctcgccg gcttgcggtg gtggcgggca cagcagcggc agtgaccact
180 cctgggctac tgagtgcgca cgttccgatg gtctccgccg aaccgtgtcc cgacgtcgag
240 gtggtgtttg cccgtggcac cggggagcca cctggtattg gcagcgtcgg aggactgttc
300 gtcgacgcac tgcgtttccc aggttggcgc caagtcactc ggggtctacg ccgttaa
357
<212> Type : DNA
<211> Length : 357
    SequenceName : gi_GDC_MTUB_4169350
    SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4169350

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggatgcat gtcattcccg ggcgcggcgc ggcgtggttg atcgtcgacg tccgagatgt
60 ggcggcactg cacgcggcgt tgttggaatc cgggcgtggg ccgcgccgct acactgcggg
120 aggtcatcgg attccggtgc ccgagctcgc gaaaattctg ggcgggtcgc cggcaccacg
180 atgctggccg tcccggtgcc cgattccgcg ctgcgtgtcg cgggatcggt gctggatcaa
240 gccgggccct atctgccttt caatactccg ttcaccgcgg caggtatgca gtactacaca
300 cagatgccgg agtccgacga ttcgccgagc gaaaaagaac taggcatcac ctaccgcgat
360 ccgcgcgaca ccgtggccga caccgtcacg gccctgcgcg gcctgggcag ctaa
414
<212> Type : DNA
<211> Length : 414
        SequenceName : gi_GDC_MTUB_4170798
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4170798

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgtgtaaag catgtctcgg tcaccatacc catcaccacc gaacatctcg gcccctacga
60 aatcgatgcc agcacgatca accccgacca gcccatcgac acggctttca cccaaaccct
120 cgatttcgcc ggcagcggca ccgtgggcgc gttcccttc ggcttcggct ggcagcagag
180 cccgggattc ttcaactcga ccacaacccc gtcgtcgggc ttcttcaact ccggcgccgg
240 tggcgcatcg ggcttcctca cgacgccgc agccgccgtg tcgggcctgg gaaacgtctt
300 caccgagact tcgggcttct tcaatgctgg cggcgtagga attcgggctt ccaaaacttc
360 ggcaacctgc tgtcgggctg ggcgaaccta ggcaataccg tctccggttt ctacaacacg
420 agcatgctgg acctcgcgac ccaagccctt atctccggct cggcaacca cggagcccga
480 ctctccggca tcctcaacaa cggtagcgga ccctaa
516
<212> Type : DNA
<211> Length : 516
        SequenceName : gi_GDC_MTUB_424142
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_424142

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgatgtgga agccgcgctg gcgatggtgt cgacggctt cggagcggcg aaccaccgcc
60 agcccagatg cctgccgcaa cgtatcgcgg tgccggtcac caagcttaag acttgccggc
120 tcgggatcac cgtggcatcg gatgcgatcg agatccacgg cggcaatggc tacatcgaga
180 cctggccggt ggcccggttg ctgcgtgacg cgcaagtcaa cacgatctgg gagggccccg
240 acaacatcct gtgtctggat gtgcggcgcg ggatcgagca gacgcgcgct cacgagacac
300 tgttggcgcg gctgcgcgat gcggtgtcgg tgtccgacga tgacgacacc acgcggctgg
360 tctcgcgccg cattgaggac ctcgacgcgg cgatcaccgc ttggaccaaa ctcgacaggc
420 agctggccga ggcgcggctg ttcccgctgg cccaattcat gggcgacgtc tacgccggcg
480 cgttgctcac cgagcaggcc gcctgggaac gggcaacccg cggcaccgac cgcaaggcac
540 tcgtcgcccg cctgtacgcg cgccggtatc tcgccgacca aggcccgctg cgcggtatcg
600 acgcagattg cgatgaggcg ctgcagcgtt cgacgaact cgtggcgggc gcgttcactg
660 ccgagcagac gtaaaagccc ccaattcgtg gctcttctga cacttccgtg ggtgagtttg
720 tgtcctgagt ag
732

```
<212> Type : DNA
<211> Length : 732
      SequenceName : gi_GDC_MTUB_4252190
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_4252190

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcgggccc cggcgacccg cgcggccagc cgcggctctt cgaggaattc cgaccagcgc
60 ccgtcgggca ggtcggtgat cccgtcgcgg ccttccagca gcgcctgcca ggtctgctcg
120 ggggtgttca tctcgcccgg gaagcgggtg acaagcccca cgatcgcgat gtcgacgcgc
180 tcggccgggc cggtgcgcga ccagtcttcg gcgtcatcgc ccgctaggtc ggtctccggc
240 tcgccctcga tgatccgggt ggccagcgat tcgatggtcg gatgcgcgaa cgccaccgcg
300 accgacagcg tgaccccggt caggtcttct atgtcggcgg ccatcgcgac ggcatcgcgc
360 gacgacagac ccagctccac catgggcacc gattcgtcga tcgagtccgg tgcctttccg
420 acggccttac ccacccagtt gcgcagccac tggcgcatct cggggaccgt tagctcggcc
480 ctttcggcgg gggcgttctc ctgggattcc gctacgtcag ccatgggtcc tcagtccgaa
540 gtggcgaaga ccgtcgggga acccacgcca ctgcgcaggc tgccgtcgag gtag
594
<212> Type : DNA
<211> Length : 594
      SequenceName : gi_GDC_MTUB_4260620
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_4260620

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcacgagg acccgcacac tggcgtcgag ccgggtgccg ttacggcgca ccgagattgc
60 cagcacccgc gcccggcctg tggcgatgag ccgttcaatc cggcgtgtgt tctcgtgcgt
120 acggacggtc ccgacgaccg gaagtgtgag atgacggcga tcaggttcga cgcgcatcgc
180 tccggtcgtg aatgtcacgc ggtcctgatc gcggcctttc ttcttgaacc gggggaagcc
240 cattgtcttg ccctcacgtt taccggatcg ggagttctgc cagttccagt acgcatcgac
300 agcgccgcca atgccgtcga cgtaagcctc tttcgagcac tccggccacc acaccgcccc
360 ggtctcggcg ttgacacaca cctcgtcctt gacggtgttc caccgtttac gaagcacccg
420 cagcgacggc ttgacagtcc cgataccagt aacgcgccac gcctcgatat cggctttcaa
480 agtagcgacc gcccagttgt aggccttgcg gcgagcgccg aaatgccgcg ccagcgcgcg
540 ggcctggtcc tcggttgggt ccagcgtgaa ccggaacgcc tgcacacacc agccttctgg
600 cacctcgaat ctggccatca agctgcctcc gcgtccccga ccgcagcagc aagggcacgc
660 ttggccccgt tctgtgcagc gcgttcacca tag
693
<212> Type : DNA
<211> Length : 693
      SequenceName : gi_GDC_MTUB_4302166
      SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_4302166

Sequence
--------
```

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcgcccgt caaggtccac cctgatagcc aaatgcgcca gctggcggca accaccccgt
60 tgtcttcgat ccgcagccgt aaaccgtcgt tcgtcggcgc ccgtcgccca acgtgaactg
120 agggcggaga atcggccgga atctcgccct cagttcacgc tcggcgccgt ttggcctcac
180 ccagtcaatg tgatctgtgc gggcggggcgt tggcgcgtag cgaacccag tggcgccggc
240 ccgccaagca cgccccggcg cggccagctc atcagcggct acgcaagcgc aacggcgccc
300 gcgatgggct gtggaagaac ccggaggatc tcaccgaaca ccagaatgcc aagctgtcgc
360 gctcatctac tcaaagaagg cctacggcac ctgttttcgg tcaaaggcga agagagtaag
420 caggcactgg accggttgat cttctag
447
<212> Type : DNA
<211> Length : 447
      SequenceName : gi_GDC_MTUB_4317863
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4317863

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcattcgg ctagctcggt tgccacaccc gtcaggggtt cgacgttggc gggttcggcg
60 ggccccagca ccgctgtcac catgcccgcc aagccgacct gcggcgccac caactgcagc
120 accagcatgt cgccgtcgcg cgccgcgatc acatggcggt cgcccctgcg gcacacgacg
180 aagcgcacca tgacgccgcc aatgtcgcgc cgccaccagc gaccctccaa ggtccgatct
240 ggcctgccca gggtttcgac catctccgcg acgtcggtt ggggctcccc gtggaggtcg
300 agcacccctt gcgctgtgag gtcacgctgc acctgttccc agacgatgtc tcgcagatcc
360 tcttgcggga tattcggccg aatcccaagc gtgacaggga aatcaaccag gtgtaaccga
420 tcggcgatca ccaacatgcc gtcgatggtt acctcgacgc cgaccacgtt gtcggcggtg
480 cccgcgcggc ctgcagcgga cggacccgtc atgatcaacc gaaaatcttg tcgataa
537
<212> Type : DNA
<211> Length : 537
      SequenceName : gi_GDC_MTUB_4341852
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4341852

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atggaccgac tctgcggtgc gccgctatgt caccgacgcc ggggccctac tgccacggct
60 gcacaagctg gtgcgcgccg actgcacgac ccgcaacaag cgccgggccg cgcggttgca
120 ggccagttac gaccggctgg aagagcggat cgcggagctg gccgcccagg aggatctgga
180 tcgggtgcgc cccgacctgg acggcaacca gatcatggcg gtgctcgaca ttccggcggg
240 cccgcaagtc ggcgaggcgt ggcgctactt gaaggagctg cggctagagc gcggcccgtt
300 gtccaccgag gaggcgacaa ccgagctgct gtcctggtgg aaatcacggg ggaaccgcta
360 gcttgggagt cgcgtcagaa cggttgtgga gtactgcata gccggcgacg acggcagcgc
420 cgggatctgg aaccgcccgt tcgacgtcga cctcgacggt ga
462
<212> Type : DNA
<211> Length : 462
      SequenceName : gi_GDC_MTUB_4391527
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_4391527

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgcttagcc tatccgctgg cggcccggaa ccgagaatgc gaccaggtca caacccagtc
60 accttccacg ccgagcagac gaggaatcgc actgcgcgga ·cctcacgcgt gcgattccgc
120 gtctgctcgt cagacaaatc agcccaggat cagcgagtcg gcgtcggggc tgacgttgac
180 cggcacggta tcgccgtcgt gcacctggcc ggccaacagc atcttggcca gctggtcacc
240 gatggcctgc tgcaccagcc ggcgcaacgg ccgcgccccg tacaccgggt cgaatccgcg
300 ctgcgccaac cagcgcttgg ccggcagcga gacctgcagc tgcagccgcc gctgcgccag
360 ccgcttgccc agctgcgcca gctggatgtc gacgatgcgc accagctctt cggggttgag
420 accctcaaag atgagcacgt cgtcgagccg gttgatgaac tccggcttga cgtagcgcg
480 caccgcggcc agcacctgct cggcgctgcc acccgacccc aggttggacg tcaggatcaa
540 gatggtgttg cggaagtcga ccgtgcggcc gtgcccgtcg gtgagccggc cctcgtcgag
600 gacctgcagc agcacgtcga acacgtccgg gtgcgccttc tcgatctcgt cgaacagcac
660 caccgtgtag ggacgccggc gcaccgcctc ggtcagctga ccgcccgcct cgtatcccac
720 atagccgggc ggggcgccga tcaaccgagc cacggtgtgc ttctcgccgt actcgctcat
780 gtcgatgcgg accatcgccc gctcgtcgtc gaacaggaag tcggccagcg ccttggccag
840 ctcggtcttg ccgacaccgg tcgggccgag gaacatgaac gccccggtgg gccggttggg
900 gtcggacacc ccggcccggc tgcgccggac cgcatcagag actgcggtaa ccgcggcctt
960 ctgcccgatg acccgcttgc ccagctcgtc ttccatgcgc agcagcttgg cggtctcgcc
1020 ttccagcagc cgaccggccg ggatgccggt ccacgccgac accacgtcgg cgatgtcgtc
1080 gggaccgacc tcctccttga gcatcacctg ctcccgggcc tgcgcctgcg gcaacgccgc
1140 gtcgagcttc ttctccacct cggggatgcg tccgtagcgc agctcggcgg ccttggccag
1200 gtcgccgtcg cgttcggccc gctcggattc cccgcgcagg gcttccagct gctccttgag
1260 gtcgcggacg atttcgatcg cgttcttctc gttctgccag cgggtggtga gctcggccaa
1320 cttctctttc tggtcggcca gctcggagcg cagcttggcc aaccgctccg ccgacgcctc
1380 gtcttcttct ttggacagcg ccatctcttc gatctccagc cggcgcacca gccgctcgac
1440 ctcgtcgatc tcgacgggcc gcgagtcgat ctccatccgc agccggctgg ccgcctcgtc
1500 gaccaggtcg atggccttgt cgggcaggaa gcgggcggtg atataccggt cgctcaaagt
1560 ggcagctgcc accagcgccg agtcggtgat gcgcaccccg tggtgcacct cgtagcggtc
1620 tttgagcccg cgcaggatgc cgatggtgtc ctccaccgac ggctcgccga cgtacacctg
1680 ttggaaacgg cgctcgagcg cggcgtcctt ctcgatgtgc ttgcggtatt cgtccagcgt
1740 ggtcgccccg accagccgta a
1761
<212> Type : DNA
<211> Length : 1761
        SequenceName : gi_GDC_MTUB_459316
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_459316

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcttgccg atttcgatgt aggacaacac cttttccagc tggtcgttgg aggcctggga
60 acccagcatg gtttcggtgt ccagcgggtc gccctgccgg accgccttgg tccggatcgc
120 cgccagctcc aggaactcgt cgtagatgtc ggcctggatc agactgcgcg acgggcaggt
180 gcacacctcg ccctggttga gggcgaacat ggtgaagcct tccagcgcct tgtcgcagaa
240 gtcgtcgtgg gcggccagca cgtcggcgaa gaagatgttg gggctcttgc cgccgagttc
300 cagggtgacc gggatcaggt tgtgcgaggc gtattgcatg atcagccgcc ccgtggtggt
360 ttccccggtg aacgcgacct ggcgatgcg gtcgctggag ccaacggct tgccggcctc
420 ggcgccgaat ccgttgacca cgttgaccac cccgggcggc aacagatcac cgatcagcga
480 catcaggtag agcaccgaag cgggtgtctg ctcggcgggt ttgagcaccg ccgtgttgcc

```
540 ggccgccaac gccggcgcca gcttccaggc cgccatcagg atggggaagt tccacggaat
600 gatctggccc accacgccga gcggctcgtg gaagtggtag gccacggtgt cctcgtcgat
660 ctggctcagc gcgccctcct gggcgcgaat cgccgcggcg aagtaccgga agtgatcgac
720 cgccaacggg atatcggcgg ccagcgcttc ccggaccggt ttcccgttgt cccagacctc
780 ggccaacgcc agcgcggcgg cgttcttgtc gatgcggtcg gcaatcatgt tgaggatcgc
840 cgcccgttcg gccggtgcgg tcttgcccca ccccggcgcc gccgcgtgcg cggcgtcgag
900 cgccttgtcg atgtcggccg cgtcggagcg cggcacctcg cagaacggct ggccggtcac
960 cggcgtcggg ttctcgaagt agcgcccatg gaccggcgcg acccactggc ccccgatgaa
1020 gttttggtac cgggattcat aggacatcag cgccccggcg gaaccgggac gggaaaagac
1080 agtcatcgta ttcggctcct cgtcaaaatc atgtaa
1116
<212> Type : DNA
<211> Length : 1116
       SequenceName : gi_GDC_MTUB_549643
       SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_549643


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgtatcttc cgcccaagct gatcccgagg cggatcccgg cgcaggtgag gccaactatg
60 gtggccccccc aagttcccca cgtcttgtcg atcacaccga atggcgcag tggggaagtc
120 tgcccagcct ccgggtctac ccgtcccaag ttgggcgtac agcctcccgc cgcctcggga
180 tggccgctgc cgacgcggcc tgggccgagg ttctcgcgct gtcaccggag gccgacactg
240 ccggcatgcg cgcgcagttc atctgccact ggcagtacgc cgaaatcaga caacccggca
300 aacccagctg gaacctcgag ccgtggcggc cggtcgtcga cgactcggag atgttggctt
360 ccggctgcaa tccgggcagc cctgaagagt cgttttagtg ctcggccaac cgactcgggc
420 gcagttggcc gcgctggtag accacaccct gctcaagcct ga
462
<212> Type : DNA
<211> Length : 462
       SequenceName : gi_GDC_MTUB_566823
       SequenceDescription :


Custom Codon
------------
Sequence Name : gi_GDC_MTUB_566823


Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgacgtcta cgaacgggcc atcggcgcgg gataccggtt ttgttgaggg ccagcaggcc
60 aagacacaac ttctcaccgt ggccgaagtg gcggccctga tgcgggtgtc caagatgacg
120 gtgtaccggc tggtgcacaa tggcgaactg cccgcggttc gggtcgggcg gtcattccgg
180 gtgcatgcca aggccgtcca cgacatgttg gagacttcgt acttcgacgc gggctag
237
<212> Type : DNA
<211> Length : 237
       SequenceName : gi_GDC_MTUB_591109
       SequenceDescription :

Custom Codon
------------
Sequence Name : gi_GDC_MTUB_591109


Sequence
```

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtggcggagt ccgtggctat ccgcggctgc ctgctgaggt gcgggccgcg ttcccgaccg
60 cggcggagat cgcgccgcag tggcatctgc gcatgcaggc cgcggtgcag cgccacgtcg
120 aggccgccgt gtccaagacg gtcaacttgc ccgccacggc gacggtcgat gacgtccgcg
180 ccatctatgt ggccgcctgg aaggcaaagg tcaagggcat cacggtgtat cgctacggca
240 gccgggaagg acaggtactg tcctacgccg cgccgaaacc gctactggcg caggctgaca
300 cggagttcag cggcggctgt gcgggccgct cctgcgagtt ctgacggcgg ctcccatggc
360 gcgagcagac gcagaatcgc acaaaatcag cgattttga
399
<212> Type : DNA
<211> Length : 399
      SequenceName : gi_GDC_MTUB_663028
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_663028

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgctgcaca gcagcttcgg gcacctcgag ggcatccagc agccgctcat agacgagctg
60 gcagaactcg accacgtgtt gggcaagctg ccggacgcct accggatcat cggccgcgcc
120 ggcggcatat acggtgactt cttcaacttc tatctgtgtg acatctcact gaaagtcaac
180 ggattacagc ctggaggtcc ggtacgcacc gtcaagttgt tcggccagcc gaccggcagg
240 tgcacaccgc aatga
255
<212> Type : DNA
<211> Length : 255
      SequenceName : gi_GDC_MTUB_688806
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_688806

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgctggggg cgctgcacca gtacccgcac actcgcatcc agccgggtgc cgttgcggcg
60 caccgtgatc gccagcaccc gcgcccggtc tttggcgatg aggcgctcga tgcggcgggt
120 gttctcatgc gtacgcacgc agccgatcac cggcaaagtg aggtgtctac ggtcgggctc
180 aacgcgcatc gcacccgtgg tgaacgacac gcgatcggcg tcgcggccct tcttcttgaa
240 tcgagggaag cccattctct tgccgtcgcg cttgccagca cgcctctgct gccagttcca
300 gtacgcgtcg accgcgcccg cgatcccgtc ggcgtaggcc tctttcgagc attccggcca
360 ccacacggtg ccagtctcgg cgttgacaca cacctcgtct ttcaccgtgt tccagcgttt
420 ccgcagtacc cgaagcgacg gcttcgccgt ctgggcgccg gtcgcgcgcc acgcttggat
480 atcggctttc agctgcgcga cggtccagtt gtaggccttg cggcgggcgc cgaaatgccg
540 cgccaacgcg tgtgcctgct cggcggtcgg atcgagtgtg aaccggaacg cttgcacaca
600 ccagccgttg gggatctcca aacgcggcat ctcaggccgc ctcatgatca tcgacagcgg
660 cagccgcgac ggcccgcttg gccgggttct gagcagcacg tttgccatac aaccttgcgc
720 acatcgaggt cagaatctcg gtcatatccc ataccaggtc atcgtcaacc tcggccgagt
780 ccaccacgac caactcccga ccctgagcgg ccagcgcagc gtggacatac tccgaaccga
840 accggcagaa ccgatcccga tgctcaacca caatccgcgt ga
882
<212> Type : DNA
<211> Length : 882

128

SequenceName : gi_GDC_MTUB_701762
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_701762

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atggcttcca gtaccgacgt gcggccgaag atcactttgg catgcgaggt gtgcaagcac
60 cgtaactaca tcaccaaaaa gaaccgccgc aacgacccgg accggctgga gctgaagaag
120 ttctgcccga attgcggcaa acaccaggcg caccgcgaga cgcggtaa
168
<212> Type : DNA
<211> Length : 168
        SequenceName : gi_GDC_MTUB_731710
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_731710

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
gtgccgccac cgatcccgcg gtgcgcggcg ccagtactt cggacccgat ggcttcggtg
60 aaatacgggg ctacccgaag gtggtggcct ccagcgccca gtctcacgac gagcagctgc
120 agcgccgcct gtgggctgtg tccgaagagc tcaccggggt cgtctatccc gtcggatgag
180 ccggactcaa cggcaacggt tggtcaacac tcgacgatgt tgactgcgac gttgatggcg
240 agcccgccgg ccgaggtttc cttgtacttg gtgtgcatgt ccgcgccggt ggcgcgcatg
300 gtgtcgatga cctggtcgag ggtgacgcga tggatgccgt cgccgcgcaa tgccatccgt
360 gcggcgttga tggccttgcc ggcggaaatc gcgttgcgtt cgatgcaggg gatctgcacc
420 agcccggcga tggggtcaca ggtcaggccg aggctgtgtt ccatggcgat ctcggcggcg
480 ttttccactt gtcgcggtgt gccgccgagg atttcagcca atccggcggc ggccatggcg
540 gccgcggagc cgacctcgcc ctga
564
<212> Type : DNA
<211> Length : 564
        SequenceName : gi_GDC_MTUB_76032
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_76032

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttggtatgcg ccgccgcccc cggtcgacga cgacccctcg gcgtaggcgg acaggtcgaa
60 gccggcacag aatccctcgc cgcgaccgga caccagaatg acatgcacgc ctggatccag
120 atcggcacgc tccaccagag cagacaactc cagcggggtg tctgcgatga tcgcgttgcc
180 cttctccggc cggttgaagg tgatccgcgc aatccgaccg gtgacctcat aggtcatcgt
240 cttcaggttg tcgaaatcga ccggcctgat cgcgtgtgtc atcagcggcc gctcagcctt
300 ttaccagcgc acgctcgagg atgggcgcga gatccagacc ggccggcatg gtgccgtacg
360 ctccgcccca ctggccgccg agccgagtgg ccagaaacgc ctcggcgacg gcgggatgtc
420 cgtggcgcac caacaacgat ccctgcaacg ccaggcagat gtcttcggca atcttgcggg
480 ctcgataacc gatcgtgtca agatcgccca gctgcggacg cagcctttcg acgtggccgt

540 ccagcctggg gtcctggcct gcgctgcggg ccagctcgtc aaacagcacc tcgacgcatg
600 cgggccgggt tgccatggcg cgcaaggtat ctagcgcgct ga
642
<212> Type : DNA
<211> Length : 642
        SequenceName : gi_GDC_MTUB_772761
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_772761

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgatcccga tggacgtgat attcggctgc ccgttgtacg ccaatttctg taagccctcg
60 gtcgtgagga agacattggg gatcttggcc agcgcggtgg aattcggcac aatgccaacg
120 acccgcaatc tgcgcgcgcc gacctcgaca gtgtcaccga ggtgtcggcc catcgtgctc
180 gatgccgcga cttcgtccgg tttcgacggt gaccgaccct ctgagacccg tggcatgcca
240 ggtccgtgct cgggcgcgcc gaagaccgtg acgtttcgcg tcgacgtgcc ttctttcatg
300 atcgtcccca cgctgcccaa cggggccgcg gccatgacac cgggttcagc ggccactcgg
360 gccaggtcaa catcgggaaa cggtattgaa cccagaaaag gtccagcagc gccggatctg
420 acgacgaata catcgacacc catggaatcg acggtgtgcc gggcctccac ccggaagccg
480 ttcgcgagtc cggtcaaaac aagcgtcatc ccgaagatca gcccggtgct gatgatcgtg
540 atgaccaggc ggcgctttct ccattgcatg tcacgcaggg ccgcgaagag cattcccaga
600 ggctaccaac gtggcgcact tgtggggcct ggtcttgacg ttttgtggtc agggcgcggc
660 ccgctagtgg tcgaagaggc gttcggggtg gtggtagtcg ttggtgtggg caccgcggtc
720 gaggtggggt ggcgggatcc attccgtttg gccgtcggac cgtttccttg tctgccagcc
780 tttcccgact ag
792
<212> Type : DNA
<211> Length : 792
        SequenceName : gi_GDC_MTUB_80423
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_80423

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgggtctcg ttgcgccggc aggtgacggt cgcgcagcga aaaagcgacc tgcgggccgc
60 cgaggatccg atcgacgccg tcgtatgcgc ctacgtggcg ttgtacgccc aacgccggcc
120 cgccgatgtc acgatctatg gggacttcac caccgggtac attgtcacgc cgtcgctgcc
180 caccgacttc agaacggcac cggacgctgg tcgacgggcg cgagcacgtc gatgaggtcg
240 accaccgtcg ccagcgcagc ggcacgcggg tcccgccctt cgaccagcgc cgagaccacc
300 gatccgtcga ccgcacagat caacgtacac accagttcga tctgtgcgga gcggccggag
360 cgctcgatgg cctcggccac ggcctcagcg cgctga
396
<212> Type : DNA
<211> Length : 396
        SequenceName : gi_GDC_MTUB_868821
        SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_868821

130

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
atgcggtgta gggcggcgtt gagctggcgg ttgcccgagc ggctgagccg catctggccg
60 gcggtgttgc ccgaccacac cgggatggga gccactgcgg catggcaggc gaaggcggct
120 tcgctttttga accgggtcac tccggcggct tcgccgacga ttttggctgc agtcagctcc
180 gcgcagccag ggatttccag cagtgcgggg gcgacctggt ggactcgggc gctgatgcgc
240 tgggctaggg tgttgatctc gccggtgagc cggatgatgt cggtcagctc ggcgcgcgcg
300 agttcggcga ccaatcctgg ctgggtgtcc agccaggtcc gcagggcctg ctggtgcttg
360 gcggcatcga gcgagcgtgc tgccggtgcc cgctcgggat cgagttcatg gacgagccag
420 cgcaaccggt tgatcgccga cgtgcgttgg gccacaagga catctcgacg gtcagtcaac
480 aacttcaact cccgcgacgt ctcgtcgtgg gtggccaggg gtaggtcggt ttcacgcatc
540 accgcccgcg ccaccgccag cgcatcgatc ggatccgact tgccccgact gcgcgccgac
600 ttgcgggtct gggccatcag cttggtgggt acccgcacca cctgctggcc ggccgccagt
660 aggtcacgct ccagacgcgc cgacatgttg cggcagtcct cgatgcccca gatcagctcg
720 aggccgaact gttcacgggc ccacatgatg gctgtggcgt gcccggccgt ggtggccttg
780 acggtcttct caccgagttg gcgacccact tcgtcggtgg ccacaaaggt gtggctgtac
840 ttgtgcgcat cggttccaac aacaaccatg gtggttgcct ctgaaccgcc ccggtga
897
<212> Type : DNA
<211> Length : 897
      SequenceName : gi_GDC_MTUB_890358
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_890358

Sequence

<213> OrganismName : Mycobacterium tuberculosis
<400> PreSequenceString :
ttgcggcgcc gagccgctgt tcctgttgga ttacatcgcc gtcggtcgga tcgtgccgga
60 gcgactcagc gcgatcgtcg ccggtatcgc cgatgggtgc atgcgtgccg gctgtgcgct
120 gcttggcggc gagaccgcag aacatccggg cctgatcgag cccgatcact acgatatctc
180 tgccaccggc gtcggcgtcg tcgaggcgga caatgtgctg ggtcccgacc gggtcaaacc
240 cggcgacgtc atcatcgcga tgggctcgtc gggtctgcat tccaatgggt actcgctggt
300 ccgcaaggtg ttgctggaga tcgaccggat gaatctggcc ggtcatgtgg aggagttcgg
360 tcgcaccttg ggcgaagagt tattggagcc gactcgcatc tacgccaaag actgtttggc
420 cttggccgcc gaaacccgtg tccggacgtt ttgccacgtc accggcggcg ggctcgccgg
480 caacctgcaa cgggtcatcc cgcatggcct catcgccgag gtcgaccgcg gcacctggac
540 acccgcgccg gtattcacca tgattgccca gcgcggccgg gtcaggcgca cagagatgga
600 gaagacgttc aacatgggtg tcggcatgat cgccgtcgtt gccccgaag acacgacgcg
660 cgccctggcc gtcctgaccg cgcggcacct ggactgctgg gtattgggaa ccgtctgcaa
720 aggcggaaaa caaggcccgc gggcaaaact ggttgggcag cacccgagat tctaagaacc
780 agacctaacc gggtctaa
798
<212> Type : DNA
<211> Length : 798
      SequenceName : gi_GDC_MTUB_904043
      SequenceDescription :

Custom Codon

Sequence Name : gi_GDC_MTUB_904043

131